Team-Based Variable Pay:

Report of the Iowa Pilot Project

Iowa Department of Education

2002

State of Iowa
Department of Education
Grimes State Office Building
Des Moines, Iowa
50319-0146

Executive Summary

The Team-Based Variable Pay Project in Iowa was designed to reward staff members for improving student achievement in their schools. In TBVP local school staff members set student achievement goals, decide on pay distribution, help students meet the goals, and are rewarded for student achievement of the goals. Eighteen schools in ten districts were accepted into the TBVP pilot project for the 2001-02 school year. The schools reflected the diversity of schools across the state.

This study describes the design and operation of the TBVP program in Iowa and in other states. It is not the purpose of this paper to make recommendations, but to review the issues surrounding TBVP. Analysis, findings, and suggested considerations about the TBVP pilot program are contained in this report. These are highlighted below. The issues are complex and multi-faceted, and are discussed in full detail in the pages that follow.

In Iowa school teams determined how much emphasis to place on the program in their school. Nine schools received TBVP awards and nine schools did not. All TBVP schools, those who received rewards and those who did not, reported progress toward their student achievement goals. Student achievement in schools participating in TBVP increased during the 2001-02 school year as measured by the Iowa Tests and local criterion-referenced assessments. The mean growth on ITBS in Reading Comprehension and Mathematics Total exceeded one-year's growth for all schools involved in TBVP. However, after only one year, the results are not conclusive.

Teachers are the most important school specific factor in influencing pupil achievement (Hanushek et al., 1998). Iowa teachers are divided in their feelings about TBVP. Key findings regarding teachers discussed in this paper include:

- Teaching is multi-dimensional. Teachers resent being measured exclusively on the test scores of their students.
- TBVP strengthened team-based cooperation among staff members in many schools.
- Teacher ownership of assessment data was increased.
- Teachers want to be seen as professionals who do not require financial incentives to increase their effort.
- TBVP increased the prominence of school achievement goals for staff.
- Communication and ownership are important aspects for effective TBVP programs.

The results of this study include many inconsistencies. Preliminary results appear that TBVP may increase focus, teamwork, and student achievement in a school building, but the cost may include teacher satisfaction and stress. Five additional highlights contained in this study are:

- Goals are motivating to teachers. Goals must be challenging, yet attainable. Variability (standard error) of assessments must be considered. Schools need help in writing quality goals.
- Assessments must be compatible with curriculum standards and teaching strategies. Teachers must have confidence that achievement gains will be measured. All assessments used for accountability must be technically adequate.
- School leadership is a crucial factor in the acceptance of TBVP and in the value of the program as seen by the staff. Communication is a critical link.

- Teachers were pleased that this was a voluntary program. Teachers indicated that they were proud that this was something that they chose to do. School personnel were able to tailor the program to fit their school.
- For an effect implementation of TVBP, negatives must be minimized. Some teachers were concerned that this is not a professional way to increase pay for teachers. A few teachers felt additional stress. Teachers were also concerned about fairness issues.

Table of Contents

Background

This study was developed on request of the legislature in Senate File 476. The intended audience consists of Department of Education staff including the Teacher Quality Team, the Legislative Education Accountability and Oversight Committee, and others interested in the quality of Iowa teachers and Iowa schools.

Questions for this Study

The overarching question for this study was "What is the effect of team-based variable pay in the Iowa schools selected for the pilot project?" In crafting an answer, the focus was on five subquestions:

1. "What is the effect of team-based variable pay on student achievement?"
2. "What is the effect of team-based variable pay on teacher quality and motivation?"
3. "What are the advantages and disadvantages of the team-based variable pay program to the pilot districts?"
4. "What benefits and disadvantages are linked to the practice of team-based variable pay in the research?"
5. "What can be learned from other states and districts who have implemented team-based variable pay?"

Rationale for this Study

Three types of contemporary performance pay proposals emerge from the research: (1) school-based performance pay plans, (2) individual-level merit pay plans, and (3) hybrid plans which include elements of both individual-level merit plans and school-or team based performance pay plans. The individual-level merit pay plans have been around for many years. The difference in the new plans is that they tend to be based on student achievement rather than subjective teacher evaluations as many programs of the past were. Contemporary performance-based pay programs tend to align with other major elements of progressive education policy including the move toward greater school accountability and standards-based reform. There is a growing list of states and schools that have implemented team-based performance plans. A number of these plans have been implemented statewide including programs in Georgia, Kentucky, Maryland, North Carolina, and Texas. Several plans are reviewed in this document.

Definitions

Detractor: phenomenon that makes it more difficult to accomplish a goal.

Educational assessments: a formal attempt to determine students' status with respect to educational variables of interest (Popham, 1999).

Enabler: something or someone that supplies the means, knowledge, opportunity, or capability to accomplish a goal.

Motivator: something or someone that supplies the incentive or a reason for doing something, that which moves to action or impels.

Professional development: according to the thesaurus of the Educational Resources Information Center (ERIC) database, professional development refers to "activities to enhance professional career growth." Such activities may include individual development, continuing education, and in-service education, as well as curriculum writing, peer collaboration, study groups, and peer coaching or mentoring.

Student performance goals: goals based on student achievement on an assessment instrument.

Team-based variable pay (TBVP): pay in addition to the base salary rewarded to a group of teachers and often other staff as the result of meeting a desired outcome. Typically the desired outcome is an improved score on a test of some kind. TBVP differs from merit pay in that all teachers benefit when a schoolwide goal is reached, rather than individual teachers receiving a bonus based on an administrator's rating.

Teacher: an individual holding a practitioner's license and who is employed, full or part-time, in a nonadministrative position as a classroom teacher, librarian, media specialist, or counselor by a school or district.

Team-Based Variable Pay in the Literature

Team-Based Variable Pay for Schools has been implemented during the last 20 years in more than 15 states. Although the programs have similar attributes, many subtle, sometimes substantial, differences can be found. Using the Educational Resources Information Center (ERIC), various web search engines, and secondary searches from article bibliographies, several primary studies have been located regarding TBVP projects in public schools. Programs for performance pay for careers other than education were not included except when the need arose to provide additional theory to clarify characteristics of TBVP in schools. There exists no shortage of opinions on TBVP or views about the designs of TBVP programs, however quality primary studies on TBVP are much fewer in number. The search of the literature base continued throughout the duration of the study.

History of Team-Based Variable Pay

Pay-for-performance was first noted in the literature in England around 1710. Teachers' salaries were based on their student's scores on examinations in reading, writing, and arithmetic. This continued until the late 1800's. Payment-for-results also made a brief appearance in Canada in 1876 causing teachers and students work harder to avoid failure (Wilms and Chapleau, 1999). The Canadian experience demonstrated that test scores could be increased quickly, as long as the subject matter could be narrowed and measured.

Throughout the twentieth century education leaders have considered rewards as a way to improve education (Stevens, Spaulding, Burleson, and Killgore, 1998). By 1918 almost half of all American schools were using performance-based pay for individual teachers. These numbers decreased in the 1920s and continued to decline until the Soviets launched Sputnik in 1957. *Equality of Education Opportunity* (the "Coleman Report") was released in 1966 with its recommendations for increasing achievement in schools. And in 1969, the idea of pay-for-performance briefly reemerged in this country as big-city schools began to desegregate. Making American schools accountable became a top priority for the Nixon administration concerned over the lack of educational achievement among the growing population of urban poor (Wilms and Chapleau, 1999). In 1983 the U.S. Department of Education published *A Nation at Risk* delineating the state of American schools and consideration of teacher pay based on student performance reemerged. Supported by this report, state governments began to implement new reward pay programs for teachers. Most of these programs focused on individual teachers. Those plans were criticized for engendering competition between teachers and failing to recognize that student performance is cumulative and dependent on the students' experience in previous classes as well as current classes (Clotfelter and Ladd, 1996). Other states began to consider TBVP. South Carolina, for example, established the School Incentive Reward Program in 1984 with reward monies used for instructional purposes by the winning schools. Performance-based school reform sought to provide stronger incentives for school achievement by linking student academic performance with financial rewards for schools and for teachers. Several states have implemented TBVP plans in the last 15 years (see Table 1). Summaries of many of these plans can be found in Appendix A. Due to recent budget concerns some of the programs including California, South Carolina, and Jefferson County, Colorado have suspended monetary awards until financial conditions improve.

Table 1
Team-Based Variable Pay Plans

State/District	Criteria	Awards	Restrictions
California (California Department of Education, 2001)	Two programs both based on meeting or exceeding growth targets or statewide annual performance targets on Stanford 9	1999-00 funded at $591 per staff FTE. Funding for the Certificated Staff Performance Reward eliminated for the 2000-01 and 2001-02 school years.	Certificated Staff Performance Incentive: awards to teachers and schools at selected low schools meeting targets; Governor's Performance Awards: awards to schools only.
Colorado, Douglas County (Hartman & Weil, 1997)	Volunteer program; school or smaller unit teams chose goals and assessments.	During 1998-99 33 of the groups that applied were successful. Participants received $413 each.	Awards to teachers.
Colorado, Jefferson County (Adsit, Carpenter, & Goff, 1998)	Volunteer program; schools chose goals and assessments.	$1000 for each certificated employee and $400 for each classified employee, funded during 1995-96 through 1997-98 only.	Awards to teachers.
Florida (Florida Department of Education, 2001).	Demonstration of sustained or significantly improved student performance.	$113 million appropriated by the 2001 Legislature. Each eligible school will receive $100 per student.	Awards to schools; may be used for staff bonuses.
Georgia (Georgia Department of Education, 2000)	Volunteer program; schools develop a plan including objectives in academic achievement, client involvement, educational programming, and resource development.	Range of $16,000 to $334,000 awarded to schools in 2000. Total rewards distributed were $12,492,000. 2001-02 award was $1900 per certified staff. 2003-04 program will be limited to 9-12 schools.	Awards to schools; may be used for staff bonuses.
Illinois	Exists in statute only. Has not been funded.		
Indiana (Huffman & Wilhelums, 2000)	Improvement in academic performance and attendance rates compared to the previous three years.	$3.2 million shared by all successful schools.	Awards to schools; may not be used for athletics or teacher bonuses.
Kentucky (Willis, Koch, Lampe, Young, Kellor, & Odden, 1999)	Increased performance on state test over a two-year period.	$27 million allocated for the biennium ending in 1998. Reward Trust Fund established to forward fund the awards.	Awards to teachers.
Maryland (Maryland State Department of Education, 2000)	State test scores and attendance rates. Schools must have two years of significant progress to be selected.	$2.75 million allocated. Mean allocation for 1999-2000 was $49,000 per school.	Awards to schools; may not be used for teacher bonuses.
Massachusetts, Boston (Boston Public Schools, 2000)	Improved student achievement including test scores, attendance rates, etc. and education reform efforts.	$500,000 awarded for 1999-2000 school year.	Awards to schools.
North Carolina (Johnson, Leak, Williamson, Kellor, Milanowski, Odden, & Hanna, 1999)	Schools that achieve or exceed expected annual gain/growth on state assessments will be rewarded.	In exemplary growth schools each certified staff member receives a $1500 bonus and each teacher assistant receives a $500 bonus.	Awards to teachers.
Oregon	A statewide system of voluntary rewards was developed but not funded.		
Pennsylvania (Pennsylvania Department of Education, 2000)	Achievement and effort as demonstrated by school improvement on the state reading and mathematics test and improved student attendance rates.	$13 million in grants awarded in 1999.	Awards to schools; part may be used for staff bonuses.

Table 1 (continued)
Team-Based Variable Pay Plans

State/District	Criteria	Awards	Restrictions
South Carolina (Consortium for Policy Research in Education, 1995; Stevens, Spaulding, Burleson, & Killgore, 1998).	Achievement criteria and attendance rates.	In existence since 1984, awards of $2500 to $72,000 were given for the 1996-97 school year. Replaced with Palmetto Gold and Silver Award schools with the Education Accountability Act of 1998. This program was not funded for 2001-2002. Next year is also questionable.	Awards to schools; may not be used for teacher bonuses.
Texas (Stevens, Spaulding, Burleson, & Killgore, 1998).	High performance standards or significant gains in performance.	Awards of $500 to $5,000 per school based on number of students in 1996-97.	Awards to schools; site based decision-making committee determine how awards will be spent.
Washington	Exists in statute only. Has not been funded.		
United Kingdom	Recognize rapid improvement or high performance.	Based on number of pupils. 200 schools will receive L5,000 (approximately $3500) and with 1,000 pupils L25,000 ($17,500).	Awards to teachers.

Measuring Student Performance

Teachers are the most important school specific factor in influencing pupil achievement (Hanushek et al., 1998). Nothing matters more than a quality teacher does. Individual teacher effects are the strongest predictors of student achievement gains followed by the student's prior achievement level and the school system effects (Sanders & Rivers, 1996; Sanders & Horn, 1998). Sanders and Rivers also note that teacher effects on student achievement are both additive and cumulative with little evidence that subsequent teachers can offset the effects of ineffective ones. Kelley (1999) examined the motivating potential of TBVP in the context of standards and assessment-based reforms. She concluded that "an important goal of education reform is to create incentives for educators to modify their skills, capacities, and teaching practices to facilitate improvement in student performance."

Creating incentives requires finding those approaches that will motivate individuals and organizations to pursue the goals of the organization. Using evidence of student achievement as part of teacher and school evaluations provides incentives for teachers to take additional responsibility for student learning. To be useful there must be evidence of individual student progress as well as aggregations of this kind of information so that changes in student populations or differences in students among classrooms do not result in unfair or counterproductive kinds of assessment (Urbanski, 1998).

Performance Elements

All states and districts found in the literature that implemented TBVP used academic achievement as at least one indicator of student performance. Achievements in reading and mathematics were most often chosen as the academic measures. Accomplishment in other areas such as science and social science were also mentioned. Some models included all taught subjects, measuring each teacher on the curriculum for his or her class. The most commonly used instruments to measure growth were paper and pencil tests. Often these tests were state tests administered at the end of the year to measure performance toward state standards.

Sometimes these tests were criterion-referenced tests designed especially to align with the state standards and benchmarks. These assessments include multiple measures such as multiple choice tests and tests with open-response items. Other states use norm-referenced standardized tests such as the Iowa Tests of Basic Skills or the Stanford 9 administered one time each year.

TVBP plans in the literature can be divided into two types (a) voluntary, often requiring an application, with locally set goals and criteria and (b) required with state or district determined goals and criteria. In Jefferson County, Colorado, for example, goals were set locally by each building each year and have included writing, language arts, reading, spelling, and science. The assessments have incorporated teacher grades, ITBS, Terra Nova, Scott Foresman/Shanker Reading Test, Hammill Test of Written Spelling, and running records improvement (Adsit et al., 1998). Similarly, in Douglas County, Colorado groups of teachers submit proposals to a review committee. The proposed goals must be beneficial to students and related to school or district objectives. Clearly stated responsibilities and timelines must also be included in the application. Past goals have included reading and writing skills, mathematics proficiency, mentorship for at-risk students, conflict management/problem-solving skills, computer/technology skills, individual learning plans, behavioral expectations, and content-specific vocabulary (Kelley, 2000). In Georgia, schools must also apply to participate. Their applications must include goals on academic achievement, client involvement, educational programming, and resource development, however the academic achievement goals are weighted heavier than the other goals (Georgia Department of Education, 2001).

States that determine goals and criteria and then financially reward schools include Florida, Indiana, Kentucky, Maryland, North Carolina, and Pennsylvania. Schools do not apply for awards, but are rewarded after all schools in the state are ranked. In Florida schools are graded on their yearly performance. An “A” performance grade and/or schools improving at least one performance grade category from the previous year are eligible for recognition and financial reward. Schools earning an “A” and/or schools improving at least two performance grade categories are eligible for increased autonomy including greater authority over the school’s total budget. Criteria to rate schools includes improvement in the school’s student achievement data (Florida Department of Education, 2001). Indiana awards their schools based on improving at least two of four areas (a) total battery score on Indiana Statewide Testing for Educational Progress-Plus (ISTEP+), (b) language arts scores on ISTEP+, (c) mathematics scores on ISTEP+, and (d) attendance rates (Huffman & Wilhelmus, 2000). The original system components in Kentucky included assessments in reading, writing, mathematics, science, and social science aligned with the state curriculum standards. An interdisciplinary component was added during the second year including assessments in arts and humanities, practical living, and vocational education (Willis et al., 1999). In North Carolina goals are set in reading, mathematics, and writing (North Carolina State Department of Education, 2000). The growth of students is determined by scores on the North Carolina End-of-Grade Tests of Reading Comprehension and Mathematics (Public Schools of North Carolina Division of Accountability Services, 2000). Pennsylvania’s School Performance Funding (SPF) is based on schools improvements in achievement as determined on the Pennsylvania System of School Assessment (PSSA). The PSSA is a standards based criterion-referenced assessment used to measure a student’s attainment of proficiency in reading, mathematics, and writing (Pennsylvania Department of Education, 2000).

Often the performance indicator also includes a measure of the percent of students tested. The concern is that some schools might choose not to test populations who for some reason might not perform well on standardized tests. For example, in Boston, Massachusetts full awards are only given if 95% of the test eligible population were tested. No award was given if less than 90% of the test eligible population was tested (Schwedel, Veysey, Conti, Kellor, & Odden, 2000). In California, 95 percent of elementary and middle schools and 90 percent of the high school students must be tested to be eligible for an award. (Just, Boese, Burkhardt, Carstens, Devine, & Gaffney, 2001). Similarly in North Carolina, K-8 schools must test at least 98% of their eligible students. High schools must test at least 95% of the students enrolled in the specific courses for which end-of-course tests are given (Johnson et al.,1999). In contrast, in Pennsylvania a school must have only an 80 percent student participation rate on the PSSA in order to qualify for a performance award (Pennsylvania Department of Education, 2000).

The use of standardized tests for accountability brings with it the entire debate that surrounds the uses and misuses of tests. Cronbach (in Linn, 1999) sums up much of the discussion, "Whenever it is critically important to master certain content, the knowledge that it will be tested produces a desirable concentration of effort. On the other hand, learning the answers to a set of questions is by no means the same as acquiring understanding of whatever topic the question represents." The bottom line is that test scores are the best measure of student achievement that we have at this time when validity, reliability, fairness, and ease of administration are considered. "Standardized tests remain the best available measures of output that are valid for comparisons over time and across schools" (Ferguson and Ladd, 1996).

Basing teacher pay on test scores has not been received well by teacher unions. By a narrow voice vote, the members of the National Education Association (NEA) at their annual meeting in Chicago in July 2000 rejected a resolution that included the conditions under which the organization would accept pay plans based on other than length of service and continued education (Archer, 2000). According to the report, any pay based on student test scores is seen as too subjective. Similarly, Burgess, Croxson, Gregg, and Propper (2001) note that one objection to a performance related pay scheme for teachers in the United Kingdom is that "teaching is multidimensional and aimed at much wider outcomes than exam results or test scores."

Menro (1998) and Wright, Horn, and Sanders (1997) refute the claim that teachers can be held accountable only for their own performance and not for the performance of their students. Menro states that effective schools have (a) achievement as their major focus, (b) expect students to achieve, and (c) have principals who do not tolerate ineffective teachers. Teacher effects are dominant factors affecting student academic gain. Furthermore, the classroom context variables of heterogeneity among students and class sizes have relatively little influence on academic gain (Wright, Horn, & Sanders, 1997).

Other indicators including staff and student attendance rates and dropout or graduation rates are also considered in many TBVP plans although these measures are not weighted as heavily as the academic measures. For instance, South Carolina's awards are based on achievement criteria along with student and staff attendance rates and student dropout rates (South Carolina

Department of Education, 1998). Criteria to rate schools in Florida include (a) statewide student achievement data, (b) student learning gains, (c) readiness for postsecondary education, (d) dropout rates, (e) attendance rates, (f) graduation rates, and (g) cohort graduation rates (Florida Department of Education, 2001). Pennsylvania currently rewards schools for effort as determined by improved attendance rates. Graduation rates will be added to the criteria for the effort award for high schools in beginning with the 2002-2003 school year (Pennsylvania Department of Education, 2000). When "softer" measures are used, strict adherence to the collection and measurement guidelines are necessary to make sure that all schools are collecting the data in the same way. For example, in Boston, Massachusetts where schools are rewarded for improved student achievement including test scores, dropout rates, and attendance rates concerns about the interpretation of the non-academic measures have arisen. The annual dropout measure included any student even if they were only enrolled in the attendance center for only a day. To combat this problem, a cohort analysis, including all students attending the school for a full year, was calculated and included to provide a more complete picture of school effect on students that were consistently part of the school's program (Boston Public Schools, 2000).

Criteria for Improvement

The criteria for improvement refers to the way the change is measured and the amount of change needed to qualify for an award. The criteria is generally measured using either cut scores or expected gain scores. With cut scores, if the school's student scores, generally measured as mean scores or percent proficient, are above the minimum score the reward is given. If the scores do not make the cut score no reward is received. A second way to set the target is as an expected gain score. The expected impact of the enablers of student improvement is considered and an expected score is set. If the students make the gains needed from one test to the next, the reward is given. The method for setting the criteria has also caused much discussion. Some states, such as Virginia, measure the performance of schools based on the percentage of students who "pass" (reach a specified benchmark on) each test. Such a system does not account for growth in individual student achievement or for differences in student background (Goldhaber, 2001). Goldhaber continues that it is likely that this type of performance indicator primarily reflects factors over which schools have no control. Under this system, schools whose students start out at a relatively low level of achievement but demonstrate significant academic gains over time, may fail to be recognized as strong performers because less that the required percentage of students reach the benchmark.

The criteria for improvement may be established using a variation of three methods (a) simple comparison with past scores, (b) comparison to a standard, and (c) improvement toward a standard. Simple comparison with past scores means that if the scores improve the school is seen as improving and meeting the goal. Different TBVP programs differ in the amount of time for which the baseline is set. For example, in Indiana's awards based on the ISEP+ exam, each school competes only against its own performance means over the previous three years. Awards are figured on a two-tier basis. Forty percent is divided among all eligible schools. The remaining 60 percent is awarded on the basis of a school's improvements in each of the four areas. To receive an award, schools must show improvement in at least two areas. Schools showing improvement in only one area are designated for non-monetary awards. Schools that rank in the top 25 percent in all four areas, whether or not they have improved, are designated in

a separate program as Four Star Schools. (Huffman & Wilhelmus, 2000). Similarly, in Maryland in order to receive monetary awards significant gains in the school performance index (SPI) must be demonstrated for from two to four years. Additionally, schools with diverse student populations also must show a significant gain among their subgroups. Schools are measured by the change in their performance over time and not by their performance compared with other schools (Maryland State Department of Education, 1999). In Boston, Massachusetts points are awarded for yearly progress based on how the school performs compared to itself from one year and to next for each indicator (Boston Public Schools, 2000). In Pennsylvania each year a school's PSSA reading and mathematics average score is compared with the school's past history on the test. An increase of 50 points or more determines whether or not a school qualifies for an award. Similarly, meaningful effort in attendance was determined to be 0.75 percent difference which determines whether or not a school qualifies for an effort award. Once a school receives an award the baseline for that school changes for the next year to the score attained to receive the award (Pennsylvania Department of Education, 2000).

Another way of measuring criteria for improvement is comparison to a standard. For example, the state sets a score above which the school will be considered to be showing excellence and if the school meets that score they are deemed to have met their goal. Often, as in the case in North Carolina's performance standards, the standard is derived from the statewide average scores. The performance standards measure the absolute achievement or the percent of students' scores in a school at or above grade level (North Carolina State Department of Education, 2000). Pennsylvania has an award called "Maintenance of High Standards" that rewards schools that have a baseline of 2850 or above and score at or above 2850 for three consecutive years, but have not received a regular achievement award (Pennsylvania Department of Education, 2000).

A third way to measure improvement would be to measure the improvement toward a standard. This takes into account the fact that different schools will be starting at different achievement levels. If they close the gap between their student's scores and the standard, they receive a reward. Mohrman, Mohrman, and Odden's (1996) analysis of TBVP encourages two levels of rewards (a) for accomplishing an improvement threshold and (b) for exceeding the school's improvement target by an additional percentage. North Carolina's growth standards, for example, are benchmarks set annually to measure a school's progress based on its previous performance, statewide average growth, and a statistical adjustment (regression to the mean) (North Carolina State Department of Education, 2000).

In order to make the rewards financially meaningful, some states have chosen to limit the number of schools that can qualify in a given year. This is accomplished by either taking a certain number or a certain percentage of the top schools. For example, the schools that scored in the top ten percent this year or the twenty schools that had the greatest gain scores. In doing this the criteria becomes a moving target. School staffs are not exactly sure how much gain they must show to outperform the others. These comparisons place more emphasis on comparisons of school to school than on performance from year to year. For example, the Mississippi system defines a "floating" index of performance referenced to the yearly average performance of students in the state. Acceptable performance is relative to the scores throughout the state in a given year. Because of this, no certain description of what level of test performance is adequate in any given year is available until the distribution of scores is completed for that year. School

staffs do not have any sense of what rate of improvement is adequate (Elmore, Abelmann, & Fuhrman, 1996). Contrast this with Kentucky's system which defines an absolute index of performance by setting a fixed proficiency level in each performance domain. The system sets a baseline measure of performance and defines a twenty-year trajectory between the baseline and the proficiency standards, then schools are judged every two years on the basis of their progress toward the fixed proficiency standard. Kentucky's system is not without it's own set of problems. The set acceptable level is absolute, a politically determined number, no one knows if it is educationally or technically feasible. Another disagreement among Kentucky policymakers concerns whether the goal should be to achieve average proficiency in each school or to achieve average proficiency for each student, a considerably more difficult task (Elmore et al., 1996).

Whatever way the targets are set, consideration should be given to the error of measurement for the administration of the assessment. Milanowski's (1999) study investigated the consistency of the performance classification of schools engaged in TBVP in Kentucky and North Carolina. He specifically looked at schools that received awards but should not have due to measurement error, and schools who did not receive awards, but should have. Systems using "alternative" methods of assessment such as student portfolios and open-ended responses were found to be more likely to have problems with measurement error than those based on standardized tests. Cross-cohort comparisons, such as comparisons of the change in scores from the fourth grade students this year when compared to last year's fourth grade students, introduce another potential source of measurement error. Population changes due to student mobility may introduce measurement error as well.

Calculation of the Change

Calculating the change can be as simple as taking the difference between two test administrations. However, this method does not take into account differences between students or schools. According to Urbanski (1998) there must be evidence of individual student progress as well as aggregations of this kind of information so that changes in student populations or differences in students among classrooms do not result in unfair or counterproductive kinds of assessment. Meyer (2000) agrees and argues that school-level average test score is a highly flawed measure of school performance for four basic reasons: (a) test score is contaminated by factors other than school performance, (b) scores tend to be grossly out of date, (c) scores tend to be highly contaminated due to student mobility, and (d) scores fail to localize schools performance to a specific classroom or grade level.

A more complex, weighted index measure is used in California, Kentucky, and Maryland. California's weighted index of student performance measures is called the Academic Performance Index (API). Each year a school has a goal to increase its API score by five percent of the point difference between its API score and the statewide target. California schools already achieving at the statewide target must make a minimum of one API growth point to achieve award status. Numerically significant student subgroups of ethnicity and socio-economic status who constitute 15 percent or more of the student enrollment or 100 students must also make 80 percent of the school's target for a school to have met its API target (Just et al., 2001). As discussed previously, Kentucky examines the change in performance in two ways, absolute performance level change and the extent to which a school has progressed toward meeting its

performance goal. Scores on the assessment are combined with other variables such as dropout rates to compute an accountability index. Each school's accountability index is compared to its baseline and improvement goal. The scores on the accountability index are reported each year, but schools are formally evaluated on the basis of change over a two year period. Each school is assigned a performance target based on past performance. Schools that exceeded their targets by a sufficient margin are given cash rewards (Willis et al., 1999). In Maryland the test scores and attendance rates are combined in a mathematical calculation that results in a school performance index (SPI). SPI is the weighted average of a school's relative distance from the satisfactory standards which measure attendance rates and student performance on the test batteries. A standards-based high school assessment program is under development (Consortium for Policy Research in Education, 1999).

Still more complex, is the use of a value-added model to account for the possible effect of child ability and family background. The value-added model uses mixed-model theory and methodology to enable a multivariate, longitudinal analysis of student achievement data. Each student's test data are accumulated over time and are linked to that student's teacher(s), school(s), and school system(s). Value-added measures require the annual use of a valid and reliable achievement test in which the items used in each test cycle are fresh, non-redundant, and tied to an underlying scale. The forms used at each grade level must include a sufficiently wide range of items such that "ceiling" and "floor" effects are highly unlikely. Scores must be reported on a common scale that spans the range of grades for which the test is appropriate (Stone, 2001). A supporter of value-added assessment, Stone lists six advantages of using value-added assessment including that it (a) expresses teacher effectiveness in terms of increase over previous performance, (b) excludes preexisting differences among students, (c) can isolate achievement effects produced by an individual teacher, (d) can account for incomplete data, and (e) permits comparisons. He believes that value-added assessment holds the promise to revolutionize education. According to Sanders and Horn (1998) "even though the driving force for the creation of the value-added model was for summative evaluation, the real power of the process lies in its ability to serve as a data source for formative evaluation and for educational research." Stone (2001) adds that "value-added assessment is revolutionary because it enables parents, taxpayers, and policymakers to see how well schools are doing without penalizing those with many disadvantaged pupils, and it enables teachers to be evaluated based on the most important factor of all, their results."

North Carolina's model is also based on a value-added concept. Two key elements of the approach in this state are: (a) focus on changes in student performance from one year to the next and (b) calculated for each student in the school. The expected growth for a cohort is derived from past statewide average growth between grades. That average is adjusted by a regression equation for ability level of the cohort and the expected effect of regression to the mean. For math and reading, "the average expected between-grade growth of each cohort on vertically equated end-of-grade tests is subtracted from the average actual growth, or the value-added by the school above that which would have been expected simply due to another year's passing…the growth indices for reading, mathematics, and writing in each grade are then standardized so that they can be added together to produce a composite index" (Johnson et al., 1999). Dallas, Texas' plan implemented in 1993 calculated awards using the current year as well as past year test scores and an adjustment for socio-economic variables also using a value-added

model. The method has been described as "incomprehensible to most participants in the process and to most outside observers…school officials neither understand the process nor have any idea what sorts of gains would have been required for them to achieve a high ranking" (Clotfelter and Ladd, 1996). Carol Ascher, Senior Research Scientist, New York University (in Olson, 1998) agrees stating that as an idea it's very appealing and it feels very progressive and fair, but the execution of it is so problematic. Olson (1998) quotes John Q. Easton, the deputy director of the Consortium on Chicago School Research as saying that he doesn't believe schools or states should use value-added exclusively. "It's just too complicated, and we see all these dozens of strange situations where you don't get a full enough picture of student achievement." Kupermintz (2002) also cites concerns about the validity of the TVAAS model. He states that using student prior achievement as a blocking factor may be a serious limitation of TVAAS due to the potential confounding of student achievement and teacher effectiveness. Kupermintz notes that the sensitivity of teacher estimates to the teachers school system context and the lack of precision of the evaluation for teachers with less student data such as those at schools with high mobility rates are also concerns. Burgess et al., (2001) agree that value-added measurement is at best a partial solution because background characteristics are still likely to be important in education performance and development at school. Many experts argue that states and districts should pay attention both to a school's absolute academic performance and to whether school factors are contributing to its students' growth.

Fairness Problems

The literature provides much discussion about ways to make the TBVP programs as fair as possible. Schools with high numbers of low socio-economic students, student mobility, limited English proficiency (LEP), or special education (SPED) students state that they cannot compete with other schools that do not have these kinds of populations. For example, in Kentucky low performing schools must make more progress that higher-achieving ones if they are to meet the state's goal for all schools. Administrators in the Jefferson County, Kentucky district which includes the Louisville metro area contend that the state's program is not fair to schools with high concentrations of poor students, special education students, or children with limited English proficiency especially when performance is improving (Jacobson, 1999). On the other hand, Sanders and Horn (1998) note that race, socioeconomic level, class size, and classroom heterogeneity are poor predictors of student academic growth. The issue is not easily mediated. One side argues that schools can fairly be held accountable only for factors that they control, and therefore that performance accountability systems should control for or equalize student socioeconomic status before they dispense rewards and penalties. Others argue that controlling for student background or prior achievement institutionalizes low expectations for poor, minority, low-achieving students (Linn, 1998).

Awards

The awards generally are either paid to the teachers, paid to a broader group of staff, or paid to the school improvement fund. In some states such as Maryland the bonus goes only to the school improvement fund and cannot be used for bonuses for teachers (Kelley, 1999). In Pennsylvania schools must divide the incentive grant using these three guidelines: (a) 50 percent must be spent on planning, delivery, and assessment of the instructional program including staff

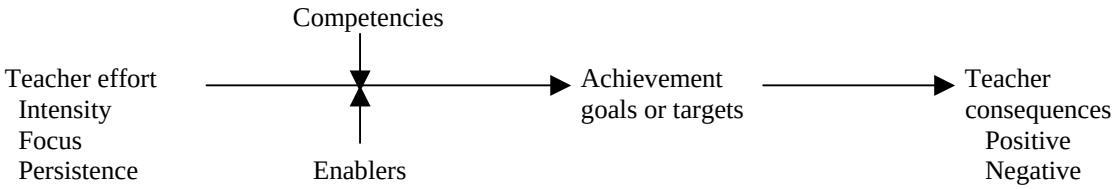
development; (b) 25 percent may be spent on staff rewards; and (c) 25 percent is at the discretion of the school committee (Pennsylvania Department of Education, 2000). Kelley (1999) found that most teachers believed that bonuses were desirable and they preferred to receive them as cash bonuses rather than have the money be contributed to a school improvement fund.

The size of the bonus also varies across states. The awards paid by the states tend to be larger when the bonuses are paid to staff rather than to the school improvement fund. As discussed in the motivation section later in this paper, bonuses must be of a magnitude that is truly noticeable and meaningful to the teachers (Heneman, 1998). The size of the reward is variable in many states dependent upon the number of units that qualify for an award. Each year in Douglas County, Colorado, a pool of money is set aside for group incentive pay and teachers meeting their goals share equally in the reward. The amount of the reward is dependent upon the number of successful teachers (Kelley, 2000). In Florida, Georgia, Maryland, and Kentucky the amount of awards varies on the number of eligible schools and the number of staff and/or students assigned to the schools selected to receive an award (Florida Department of Education, 2001; Georgia Department of Education, 2001; Maryland State Department of Education, 1999; Willis et al., 1999). In Pennsylvania monetary awards are based on the degree of improvement as well as the number of students enrolled and the number of schools that qualify (Pennsylvania Department of Education, 2000).

Enablers for the Programs

The goal of a pay-for-performance system is not just to reward teachers appropriately, but to motivate higher performance. Odden, Kellor, Heneman, and Milanowki (1999) believe that enabling conditions contribute to a school's likelihood of success in a TBVP program. They note that enablers in TBVP schools are of three types: (a) teacher knowledge, skills, and abilities, (b) tools and resources, and (c) organizational conditions. The identification of all three types of enablers are a critical factor in a TBVP program.

Teachers need to have a clear understanding of their strengths, weaknesses, and opportunities to remedy their shortcomings. Availability of professional development aligned with the goals of the pay-for-performance system becomes particularly important (Education Commission of the States, 2001). Professional development is the enabler with the greatest potential for helping teachers improve their knowledge, skills, and abilities (Heneman, 1998). Heneman studied 12 schools from the Charlotte-Mecklenburg School District including four high schools, four middle schools, and four elementary schools. In each group two two-year bonus winners, one one-year bonus winner, and one nonwinner was included. He concluded that professional development provided teachers should seek to raise their perceptions of their own competencies and of the presence and supportiveness of various enablers. He proposed this theoretical framework based on expectancy theory:



A district must plan on undertaking specific actions that support teachers in their striving to meet the goals and remove unnecessary roadblocks to goal attainment. These must be built into the original design of the TBVP program and either accompany or precede its full implementation so that teachers are likely to experience initial success in goal attainment. These initial successes and the resultant high expectancies should carry forward to goals in future time periods. "Initial success will help teachers believe that future, higher goals are also attainable if they work hard for them" (Heneman, 1998).

Both the quality and duration of the professional development program are critical to improving teaching and student achievement. Schalock (1998) notes that the aim of any accountability system should be to develop professional responsibility rather than to develop an external hammer for performance. Joyce and Showers (1995) list eight essential attributes necessary to accomplish increases in student achievement through professional development:

1. The focus must be on instruction and curriculum.
2. The study of implementation is built into the plan.
3. All site personnel responsible for instruction participate in the professional development.
4. Specific student learning goals direct the professional development efforts.
5. Intensive professional development is provided including theory, modeling, and opportunities to practice.
6. Collaboration opportunities for teachers occur on a regular basis.
7. Ongoing follow-up and support are available.
8. Evaluation is based on data related to the goals such as student progress and implementation data.

Menro (1998) suggests a differentiated staff development policy to allow the effective teacher more freedom to pursue individual interests and require the ineffective teacher to target particularly ineffective practices.

The second type of enabler is tools and resources. These relate to instructional methods and techniques (Odden et al., 1999) such as an articulated and aligned curriculum and access by teachers to instructional strategies including teaming and student grouping. Resources are used to provide quality professional development.

The third type of enabler has to do with organizational conditions. Examples include principal leadership, school-based management, district support, community support, and program alignment. An effective principal who is the building instructional leader is essential to improving student performance (Supovitz & Poglinco, 2001). They discussed three indicators seen when the principal is the instructional leader:

1. The principals are focused on instructional improvement and have a vision of instructional quality.

2. The principals developed a community of instruction practice. Collaboration and communication are emphasized in an environment that fosters improved instruction among teachers.
3. The principals rearranged priorities to focus on academic performance and instructional improvement. This does not mean that the managerial and political aspects of their jobs are ignored, just that instructional improvement takes priority.

Kelley, Heneman, and Milanowski (2000) suggest that motivational impact of a TBVP program is enhanced with active support from principals. Supportive principals generate continuous interest and enthusiasm for a TBVP program. In their study of programs in Charlotte-Mecklenburg and in Kentucky, they found a high level of variability in the extent to which the principals promoted the program. They noted that "while some principals were very proactive, others seemed to be genuinely at a loss as to how to rally their staff around the goals."

Student Achievement

Empirical studies examining the effect of TBVP on student achievement are scarce in the literature. Part of the problem is that often TBVP is often only one part of a multiple faceted effort to improve student achievement. North Carolina, for example, has shown considerable gains on the National Assessment of Education Progress (NAEP) during the last decade in grades four and eight math (Triplett, 1997; Nation's Report Card, 2000; Grissmer & Flanagan, 1998). For example, in 1992, 50 percent of the fourth grade students were estimated to be at or above the basic level as indicated by the results of the NAEP. By 2000, 76 percent of the fourth grade students were estimated to be at or above the basic level. Gains in the percent of students at or above the proficient level were just as pronounced, from 13 percent to 28 percent in grade four and from 9 percent to 30 percent in grade eight. Gains in reading have also been seen although not so extreme. The fourth grade students moved from 56 percent at or above the basic level in 1992 to 62 percent in 1998. No longitudinal data is available for eighth grade reading. During that same time period North Carolina educators have implemented a variety of strategies including redesigning mathematics standards, requiring Algebra I for high school graduation, strengthening teacher preparation, and during the 1996-97 school year, TBVP.

Another state with a long history of TBVP is Kentucky. Large gains in scores on the Kentucky Instructional Results Information System (KIRIS) assessment were reported when the Kentucky Education Reform Act (KERA) results were released for the first biennial period. Proponents of the KERA program touted the gains as an indication of the success of the program. A study by Koretz and Barron (1998) was undertaken to explore the validity of the claims. Koretz and Barron's hypothesized that the observed gains were due to one of three components: (a) meaningful gains that should generalize to other similar assessments; (b) meaningful gains that may not generalize well to NAEP or other similar assessments due to idiosyncratic aspects of KIRIS; and (c) inflation of KIRIS scores. Their study examined both internal evidence and external evidence of the legitimacy of the gain scores. Kentucky had also shown growth on the NAEP during the last decade although not as pronounced as in North Carolina (Nation's Report Card, 2000). Kentucky fourth grade students moved from 51 percent at or above basic in 1992 to 60 percent at or above basic in 2000 in math. Similar results were seen for reading, 58 percent in 1990 to 63 percent in 2000, and for eighth grade math, 43 percent to 63 percent. No longitudinal data is available for eighth grade reading. Although these increases were seen in NAEP scores, they were not of the magnitude of the gains seen in the KIRIS scores (Hambleton

et al., 1995; Koretz & Barron, 1998). For example, according to Koretz and Barron, while the NAEP mathematics scores of Kentucky students increased a considerable amount between 1992 and 1996 (0.17 standard deviations in the fourth grade and 0.13 standard deviations in the eighth grade). The gains on KIRIS were much larger, roughly 3.6 times as large in the fourth grade and 4.1 times as large in the eighth grade. Koretz and Barron charge that the disparity between the two trends suggests that the KIRIS gains were considerably inflated. They further note that Kentucky's increases on the NAEP may have simply been part of a broad national trend as the increases are similar to the national average. To further complicate the matter, TBVP was not the only implementation in Kentucky during this time. A movement to a school-based performance assessment system aligned with the state curriculum standards has also taken place during the past decade.

Koretz and Barron (1998) suggest a couple of additional possibilities for the inflation of the KIRIS scores. Evidence of a "sawtooth" pattern was seen in mathematics results in which item performance increased when items were reused but dropped again when new items were introduced. This hints that some teachers may be inappropriately focusing on reused test items. Results from reading were less consistent. Since 1993, over 60 Kentucky schools have been investigated for allegations of cheating by teachers or administrators (Becker, 1998). However, no mention was made of how many of the allegations were confirmed. Another possibility (Koretz & Barron, 1998) is that the large initial gains stem from familiarization to the test instrument. Even though the scores after familiarization to the instrument may be more valid, this still causes the gain scores to be inflated. The study concludes that the four year period of the study was too short to sufficiently study inflation due to familiarity.

Clotfelter and Ladd (1996) compared TBVP programs in the state of South Carolina and the Dallas Independent School District. However, when they focused on student outcomes, they were only able to use the data from Texas because the school incentive program in South Carolina was introduced as part of a comprehensive package of school reform making isolation of the effects of the incentive program on student outcomes in that state very difficult. The accountability and school incentive system was the centerpiece of the school reform effort in Dallas. Clotfelter and Ladd compared the student performance in Dallas on the Texas Assessment of Academic Skills (TAAS) with student performance in five other large Texas cities including Austin, El Paso, Fort Worth, San Antonio, and Houston for the school years 1990-91 to 1993-94. Focus was on reading and mathematics results for seventh graders. They found that the Dallas program has resulted in some positive effects on student outcomes since 1991. The proportion of these gains attributed to the accountability program was difficult to assess.

Ladd (1999) extended this study. She states that the relevant analytical question is not simply whether student performance on the TAAS improved in Dallas during the 1990s, but whether it improved relative to what would have been predicted in the absence of the program. She used panel data techniques to compare average student performance by school in Dallas with the five other Texas cities. Studying seventh grade during the period from 1990-91 to 1994-95, she found that Hispanic and white children in Dallas exhibited constantly positive and relatively large gains relative to the other cities, but African American students did not. She examined third grade scores and found no difference. However, she noted that the third grade results should be interpreted cautiously as there were considerable limitations including a change to a

more difficult test for third graders during the last two years of the study. Her study also found that the Dallas dropout rate had fallen in comparison to the other cities and principal turn over rate had increased compared to the past. Both which indicate, according to Ladd, a real change in the system that could suggest positive outcomes for the future.

During the same time period as Ladd's study, Grissmer and Flanagan (1998) explored the rapid achievement gains in North Carolina and Texas. During the period from 1990 to 1997, North Carolina and Texas posted the largest average gains in student scores on the NAEP. The study concluded that the most plausible explanation for the test score gains are found in the policy environment established in each state including leadership from the business community, political leadership, and consistency of the reform agenda. The reform agenda at that time included state-wide academic standards, holding all students to the same standards, computerized feedback systems, and accountability systems with consequences for results including TBVP.

Cooper and Cohn (1997) completed a comprehensive study of student achievement using data from 541 classrooms in South Carolina. Their study controlled for the numerous factors that affect the educational performance of individual students measured as gain scores on reading and mathematics standardized tests. An assortment of independent variables were incorporated including two teacher incentive plans designed to monetarily reward teachers and/or schools that are effective in increasing student test scores. PLAN1 was an individual bonus model that included attendance, a performance evaluation, and completion of self-improvement goals. PLAN2 was the TBVP design. Other variables considered included family background and demographic variables of the students, teacher variables, school resources, and innate ability of the students. Both ordinary least squares (OLS) regressions and an estimation of a frontier production function were calculated. They found that the only consistently significant variables in the achievement gain process were the two teacher incentive plans and the percentage of students eligible for free or reduced-fee lunch (PCTFRL). Their conclusion was that South Carolina leaders should consider improving socioeconomic conditions and reallocating resources in a more efficient manner in order to improve standardized test results.

The Georgia Department of Education (2000) found that during the 1996-97, 1997-98, and 1998-99 cycles, mean ITBS Total Reading and Total Math test scores of schools participating Georgia's TBVP program were significantly higher than scores of other schools in the state at grades three, five, and eight. The demographics of the TBVP schools were similar to the state as a whole during those years. Average student scores of schools involved in TBVP in 1993-94, 1994-95, and 1995-96 were also higher than other schools in the state, but the number of schools was too small and unrepresentative to make meaningful comparisons. The number of high schools receiving awards in any year was too small to make meaningful comparisons.

Teacher Motivation

One of the major goals of Iowa Senate File 476, the Teacher Quality Legislation, was to create incentives for teachers to improve their skills and teaching practices to enhance student achievement. Milanowski's (1999) study asked "What aspects of a school's capacity to improve student achievement are likely to change substantially from cycle to cycle? Capacity is a combination of staff motivation and ability. Ability is likely to change rather slowly."

Therefore, perhaps motivation is the key to improvement of student achievement. This section discusses the research on using TBVP to motivate teachers to modify their work to improve student performance. Several studies have examined the affect of team-based performance pay on teacher motivation.

Hajnal and Dibski (1993) discuss the motivational effect of compensation in TBVP compared to individual merit pay. According to Hills (in Hajnal & Dibski, 1993), the "objectives of compensation are to get employees to join the organization and to motivate them to behave in ways that are congruent with the organization's needs." Hajnal and Dibski suggest that a better understanding of the reward system may be accomplished by considering a tripartite compensation system consisting of (a) pecuniary rewards, (b) non-pecuniary extrinsic rewards, and (c) non-pecuniary intrinsic awards. Pecuniary rewards include the base salary and merit pay. Non-pecuniary rewards are those bestowed by others in the school environment. Principal feedback such as formal or informal recognition, favorable job assignment, and involvement in decision-making could be considered non-pecuniary external rewards. Hills (in Hajnal & Dibski, 1993) notes that the potential for non-pecuniary extrinsic rewards in education is immense. In addition to the principal rewards, parents, students, other teachers, family, school board members, and society dispense extrinsic rewards to teachers. For Hills non-pecuniary intrinsic rewards include status, autonomy, and a sense of accomplishment. Feelings of being appreciated, completing an assignment well, being part of a successful group, and having contributed to the success of others can all be considered intrinsic rewards. The literature is filled with studies affirming the central role played by intrinsic motivation in facilitating professional development of teachers (Feistritzer, 1986; Bookhart & Freeman, 1992; Espinet, Simmons & Atwater, 1992; Serow, 1994; Rogers, Bond, & Nottingham, 1997).

Hajnal and Dibski (1993) believe that TBVP based on school-level increases in scores encourages co-operative behavior among teachers including improved communication and joint problem solving which have been problems with individual merit pay. While they state that TBVP does not have motivating power to stand alone, it would complement a career ladder program. They emphasize that the rewards, pecuniary and non-pecuniary, intrinsic and extrinsic, must fit the tasks to be accomplished, the culture of the school, and the professionalization of teaching. For example, if teachers are expected to collaborate as a group, then somehow their rewards must reflect and reinforce teamwork.

Clark and Wilson (1961) developed a similar typology. They differentiate among three types of incentives (a) material incentives: tangible rewards including wages and merit pay; (b) solidary incentives: intangible rewards derived from association including sociability, status, and identification; and (c) purposive incentives: intangible rewards related to the goals of the organization including satisfaction from making a difference. They suggest that although organizations may include all kinds of incentives often one is predominate. For example, organizations that rely mostly on material incentives include business firms, trade unions, and political machines. These organizations explicitly seek material rewards for their members and develop precise cost-accounting machinery. Executives devote their energies toward obtaining material resources. Solidary organizations include service-oriented voluntary associations and social clubs. Efforts are directed toward obtaining prestige or fellowship. Purposive organizations involve members because they want to help achieve the goals of the organization.

Schools often embody the strengths and weaknesses of the purposive incentive organizations. According to Scott (1987), schools "must supplement their primary incentives with material inducements – to retain a core staff – and solidary incentives – to help sustain the rank and file through the dry spells."

Koehler (1996) notes that often young faculty members have prevailing financial needs including the repayment of student loans, growing family needs, and other personal commitments. He believes that financial motivations are strongest during the early and middle-career intervals. Once those needs are satisfied, intrinsic rewards and status-recognition rewards become the prime motivators for most faculty members.

Kelley (1999) studied the motivational impact of TBVP programs in Kentucky, Charlotte-Macklenburg, North Carolina, Douglas County, Colorado, and Maryland primarily through the use of interviews and surveys. She structured her study using three different theories of motivation from the employee motivation literature: expectancy theory, goal-setting theory, and systems theory. Vroom's (1964) expectancy theory suggests that the individual will consider the outcomes associated with various levels of performance (from an entire spectrum of performance possibilities) and elect to pursue the level that generates the greatest reward for him or her. Vroom's theory states that individuals will be motivated to perform based the belief that effort exerted will lead to the desired performance and the subjective probability that a given performance will lead to certain desired outcomes. In other words, "teachers must know and understand the goals; they must believe that the accomplishment of these goals is substantially within their control; and they must believe that achievement of the goals will result in an outcome (or a set of outcomes) that is valuable to them" (Kelley, 1999). Valued outcomes for teachers include professional efficacy, professional collaboration, and financial incentives. Goal-setting theory (Locke, 1968) suggests that goals that lead to high levels of performance are specific, difficult, attainable, and worthwhile. Kelley (1999) notes that research has shown that simply "setting clear and measurable goals can motivate employees to higher performance." Goal success can lead to a sense of achievement as well as to external rewards such as recognition. Systems theory is basically concerned with problems of relationships, structures, and interdependence. All systems are characterized by a combination of parts whose relationships make them, to some degree, interdependent (Hong, Al-Khatib, Magagna, McLoughlin, & Coe, 2001). Interactions among the parts of a system must be recognized and understood or change will fail. Scott (1992) advocates that the alignment of organizational resources and policies is motivating.

Kelley (1999) examined the extent to which TBVP programs motivate teachers by providing valued outcomes, clear goals, and resource alignment. Her research suggests an interactive effect, "the use of extrinsic rewards, in the form of school-based performance awards, can motivate some individuals directly and can also create intrinsic rewards that are likely to enhance teaching practice (e.g., clear goals, opportunities for professional development, and collaboration around curriculum and instruction.)." Kelley found that performance bonuses had strong direct or indirect effects on motivation to change teaching practice. On her survey, 74 percent of Kentucky teachers and 92 per cent of Charlotte-Mecklenburg teachers indicated that receiving a pay bonus was either desirable or very desirable. She found that the rewards teachers indicate that they value most were intrinsic rewards. Satisfaction from student performance improvement,

meeting school goals, and the opportunity to work collaboratively with other staff members on curriculum and instruction were the most likely intrinsic outcomes of a TBVP program.

Kelley (1999) found that TBVP could also create some negative consequences including stress on teachers, an overly focused curriculum, and pressure on assessment instruments. She suggests that many of these negative effects can be avoided through careful program design. One way to limit the negative effects is by structuring the programs to reward improvements in performance in order to control for past variations in achievement levels among schools. She calls for further research to compare and understand the differences in motivational impact of TBVP programs that provide funds for school improvement and those that provide salary bonuses.

Heneman published two articles on the motivational affect of TBVP (Heneman, 1998; Heneman & Milanowski, 1999) that studied the Charlotte-Mecklenburg School (CMS) TBVP and complemented and extended Kelley's work. His speculation, based on expectancy theory, is that, "the more that a teacher perceives that working hard will pay off in higher student achievement, the more the teacher believes that effort is accurately focused on the student achievements desired by the school, and the more the teacher is willing to persist in the pursuit of student achievement, the greater the teacher's expectancy that individual effort will likely have a positive impact on student achievement." In the first study, Heneman's team interviewed administrators and teachers in 12 schools in the Charlotte-Mecklenburg District. He found that teachers felt that the student achievement goals were understandable, challenging, specific, and focused. Many teachers stated that they believed they had a reasonable chance of meeting this year's goals. Teachers believed that their schools had adequate available resources, but they were divided on the impact of other enablers including curriculum alignment, team teaching and planning, best-practice information, professional development, and parental support. Many teachers were critical of the bonus size. Teachers reported that the largest positive consequence was helping their students learn. Teachers also noted that meeting the achievement goals brought professional pride and valued public recognition. While teachers certainly appreciate more money, there is little evidence that it drives their behavior in the classroom. They seem more likely to be gratified by "psychic rewards" such as small, immeasurable signs of individual student progress (Lashway, 2001). High stress, high pressure, and more working hours were also reported.

Heneman also compared his results with the Kentucky TBVP plan studied by Kelley and Protsik (1997). Teachers in both districts agreed that it was appropriate for teachers to be held accountable for student achievement. In both groups it appeared to be the "presence and challenge of the goals and accompanying pride in meeting them…that were the primary incentives for enhanced teacher motivation." The enabler with the greatest potential for helping teachers improve was professional development. Several differences in the CMS plan and the Kentucky plan also immerged. The measures of student achievement are more performance-based and complex in Kentucky than in CMS. The size of the bonus in Kentucky was between $1,300 and $2,300 per teacher and teachers generally felt that this was acceptable. The CMS teachers were critical of their bonus amount. Kentucky teachers were acutely aware of negative sanctions their schools could experience including state management of schools and dismissal of teachers and these negative sanctions appeared to be motivating. In CMS, negative sanctions

were either unknown or a remote possibility for the teachers. Heneman concluded that the CMS program was effective in focusing and channeling teacher effort toward improved student achievement, however, the full potential of the program was not realized due to teachers' perceived deficiencies in knowledge about various factors including the amount of the bonus.

In the second study (Heneman & Milanowski, 1999), surveys were administered to teachers in Charlotte-Mecklenburg and in Kentucky. The surveys were designed to measure the desirability of the outcomes of TBVP including bonus fairness, bonus value, pay satisfaction, bonus motivation, bonus continuation, and withdrawal intentions. The two groups were very comparable in their results. Very desirable outcomes included: personal satisfaction from improved student performance, having students learn new skills, receiving a bonus, working toward clear school wide goals, personal satisfaction in meeting goals, receiving funds for school improvement, working cooperatively, and public recognition. Several outcomes were also rated as very undesirable including more pressure and job stress, public criticism or embarrassment due to not achieving goals, loss of professional pride, putting in more hours, and less freedom to teach things unrelated to goals. TBVP is more than just a bonus program. TBVP programs trigger and deliver multiple outcomes to teachers, some of which are desirable and some of which are objectionable. The authors concluded that bonuses are very desirable by teachers and in TBVP bonuses are accompanied by a variety of other extrinsic and intrinsic outcomes associated with attaining student achievement goals. However, two major drawbacks, undesirable outcomes associated with sanctions and stress and problems in fairly administering the TBVP may cause teachers to desire to see the program discontinued or to withdraw from participation in it. Heneman and Milanowski call for careful planning, design, and administration of a TBVP program in order to realize its high motivational potential. In her review of Milanowski's work Malen (1999) notes that motivation is a complex phenomenon and the reactions vary considerably. Malen further suggests that perhaps different kinds of incentive systems are needed for schools to take advantage of the many motivational factors that inspire different staff members. She argues that the optimism of Kelley and of Heneman and Milanoski regarding the motivational potential of TBVP may be due to the newness of the program and may not pan out over the long haul.

Kelley and Protsick's (1997) study centered on six award winning Kentucky schools, four elementary and two middle schools, in the Louisville-Frankfort-Lexington areas and focused on employee motivation and compensation design. Their findings were that, rather than the monetary award, the teachers were more motivated by other aspects of the accountability program including "the fear of sanctions; the desire for positive public recognition; having clear goals, clear technologies for achieving those goals, and measurable results; and by sharing with their students the energy and enthusiasm that comes from engaging together in learning."

Group incentive programs have been found to be a component that encourages cooperative efforts within schools to work on common student goals that impact student performance. Burgess et al. (2001) note that one objection to an individual performance relate pay scheme for teachers is that teaching involves team-based cooperation that is inconsistent with an individual merit pay scheme. TBVP overcomes this objection. Others argue that just because effort is collaborative, it does not necessarily follow that all pay should be distributed equally. Some members of the group may be better players than others and advance the goals of the team better

than others. In her review of Milanowski's work Malen (1999) argues that it may be unrealistic to expect a modest school-based award to engender collegial interaction or foster coordinated collective action.

Kelley, Odden, Milanowski, and Heneman (2000) studied three sites, Charlotte-Mecklenburg, Kentucky, and Maryland, through on-site interviews with teachers and principals and mail surveys. They utilized two theories of individual motivation, expectancy theory and goal-setting theory. Expectancy theory supports that individuals are most motivated when they have a strong belief that they can achieve specified goals. Goal-setting theory states that clear and specific student achievement goals are more motivating for teachers than unclear or conflicting goals. Kelley et al. propose six conditions that must be present to maximize the likelihood that SBPA programs would have positive impacts on teachers.

1. Teachers must believe that if they try they can succeed in achieving goals.
2. Positive outcomes associated with the program must be greater than the negative outcomes.
3. Bonuses must be aligned with other motivating outcomes (such as seeing one's students achieve at higher levels).
4. Program goals must be consistent with the goals of other improvement programs in place at the schools.
5. The program must be perceived as fair, both in likelihood of success and in its operation.
6. The program must be properly implemented.

Again teachers reported that a variety of positive and negative outcomes were associated with TBVP programs. Valued outcomes were both intrinsic and extrinsic motivations including receiving a bonus, receiving school improvement funds, personal satisfaction for meeting goals, personal satisfaction from improved student performance, and avoiding embarrassment due to not achieving goals. Schools with higher average levels of teacher expectancy were more likely to show improvement in student achievement outcomes. This finding is consistent with motivational theory and earlier research on work motivation (Odden et al.,1999). Odden et al. notes that the motivational effects of the programs were not as strong as might be expected. He concludes that a successful TBVP program involves more than promising school staff a bonus if performance goals are met. Attention must be paid to design, management, and enabling conditions. Of special importance is paying attention to building teacher expectancy, ensuring the fairness of the program, guaranteeing the availability of bonuses if the goals are met, and ascertaining active support from building leaders.

While "incentive" and "bonus" are both used throughout this paper. A distinction is made between the two in pay for performance programs discussed in the literature. Bonuses are reinforcements offered after the fact. The theory is that behaviors occur due to past experiences with reinforcement and the objective measurement of the value of past rewards. Incentives, on the other hand, cause staff members to adjust behaviors due to anticipation and subjective weighting of future rewards. With an incentive program the teachers know the criteria needed to receive a reward before the school year begins. The setting of the criteria defines whether the TBVP will serve as a bonus or an incentive depending on which psychological theory is employed. No studies were found that discussed the differences between incentive and bonus programs in TBVP.

Additional Benefits and Concerns

A few additional benefits and concerns about TBVP are sprinkled in the literature (Table 2). Cooper (2002) notes that TBVP encourages team working and co-operative behavior. He also found that TBVP is an effective way to clarify goals and priorities within a team. Wilms and Chapleu (1999) state that inevitably children wind up the losers because curricula are narrowed to include subjects that can be taught by drill and repetition and that are easily measured. Teachers in the Colonial School District in Pennsylvania reported pressure from some teachers and union leaders to any teacher receiving an award to donate it to a charitable institution. The message was clear that they do not deserve the money even though they earned it (Sultanik, 2000). There is a danger that peer pressure will intensify causing an oppressive atmosphere (Cooper, 2002).

Unfortunately, there is virtually no information on how the incentives resulting from accountability systems affect other student outcomes. For instance, TBVP may influence dropout rates or the classification of students into various at-risk categories (Goldhaber, 2001). "Rewards do not provide the kind of silver bullet that will transform attitudes or jump-start a dysfunctional school. Boosting student achievement requires a comprehensive approach that includes teacher development, adequate resources, and organizational support. Done well, incentives can be a useful supporting role; done carelessly, they can create dissension that diverts attention from the central goal of improving student achievement" (Johnson et al, 1999).

Table 2
Some Advantages and Disadvantages of Team-Based Variable Pay Programs

Advantages	Disadvantages
Provides clear focus for a building to focus efforts. (Clotfelter & Ladd, 1996; Kelley, Milanowski, & Heneman, 2000)	Focuses too narrowly on the measures that are used in the rating system.
Encourages teachers to support and learn from each other	Teachers who do not contribute to the success of the building share in the reward.
Teachers see salary bonuses as being desirable. (Heneman & Milanowski, 1999; Kelley, Milanowski, & Heneman, 2000)	The greater the incentives, the more likely that they will lead to stress, strains, and distortions. (Clotfelter & Ladd, 1996)
Non-monetary rewards often accompany salary bonuses.	Non-monetary rewards are stronger motivators for teachers.

The Iowa Pilot Project Overview

Iowa Senate File 476 provided for a pilot program for TBVP for Student Achievement. The Department of Education was charged with establishing a pilot program. Schools accepted in the pilot program were required to administer valid and reliable standardized assessments at the beginning and the end of the school year. If the attendance center demonstrated growth in student achievement all licensed practitioners employed at the center would share in a cash award. Each participating school designed its own program including (a) student performance goals, (b) student performance levels, (c) multiple indicators to determine progress, and (d) a system for providing the financial rewards. The plans were approved by each local school board.

The application form is found in Appendix C. Plans were accepted at the Department of Education until October 1, 2001.

Once the plans were received at the department, they were checked for accuracy and completeness by two member teams composed of Department of Education Consultants. The teams completed checklists by assigning points for assessment information, appropriate goals set, local board approval, and description of readiness. The schools receiving the highest total points were accepted into the pilot program. Schools accepted into the program were notified on October 24, 2001. The remainder of this study focuses on the schools accepted into this pilot program.

Method

To investigate these questions, primarily qualitative methods were used. Since this study sought to identify how school districts orchestrated their pilot projects to achieve gains, it employed in-depth interviewing, participant observation, and content review to understand how the component parts fit together to create a culture for success in meeting student achievement goals. Quantitative methods were used to augment the study. For example, quantitative methods were used to compare baseline and end-of-pilot assessment data.

Participants

Thirty-one schools representing fifteen districts submitted proposals to become pilot schools. Eighteen schools in ten districts were accepted into the pilot project based on their fulfillment of the required criteria as recorded in their team-based variable pay applications. The schools are diverse in geography, student demographics, and student achievement (see Table 3).

Table 3
Schools Selected for the Pilot Project*

District, School	Grade levels	Student Enrollment	Certified Staff FTE	Enroll/ FTE	%Receiving Free or Reduced Lunch	%Minority	Population of city or town (2000)
Davis County, Davis ES	PK-4	476	39.37	12.1	37%	3%	2,601
Davis County, Davis Middle	5-8	376	24.05	15.6	32%	2%	2,601
Des Moines, Oak Park ES	K-5	411	32.4	12.7	57%	26%	198,682
Griswold, Elliott ES	K-5	133	10	13.3	31%	3%	1,039
Johnston, High School	9-12	1291	75.4	17.1	3%	7%	8,649
Johnston, Lawson ES	K-5	444	36.6	12.1	7%	8%	8,649
Johnston, Middle School	6-8	1043	75.3	13.9	3%	7%	8,649
Linn-Mar, Indian Creek ES	K-5	430	28.5	15.1	13%	5%	26,294
Missouri Valley, Middle School	6-8	243	16.48	14.7	27%	3%	2,992
NE Hamilton, K-12	K-12	295	26	11.3	20%	4%	235
Oelwein, Harlan ES	K-5	130	9.28	14.0	68%	10%	6,692

Table 3 (continued)
Schools Selected for the Pilot Project*

District, School	Grade levels	Student Enrollment	Certified Staff FTE	Enroll/ FTE	%Receiving Free or Reduced Lunch	%Minority	Population of city or town (2000)
Oelwein, High School	9-12	478	36.36	13.1	23%	1%	6,692
Oelwein, Middle School	6-9	381	30.2	12.6	39%	3%	6,692
Oelwein, Parkside ES	K-5	112	9.55	11.7	63%	12%	6,692
Oelwein, Wings Pk ES	PK-5	394	30.3	13.0	39%	4%	6,692
Stratford, ES	PK-6	99	10.6	9.3	25%	0%	746
Van Buren, Douds ES	K-6	156	13	12.0	30%	2%	unincorp
Van Buren, Stockport ES	K-6	168	13.5	12.4	33%	1%	284

*School demographic information based on the Fall 2001 BEDS documentation.

Data Collection

Interviews

Primary data collection took place over the period lasting from November 29, 2001 through July 31, 2002. During that time period, two rounds of semi-structured contacts were conducted with an administrator in each building. A semi-structured interview protocol was developed to guide the interviews. The first round of administrator in-depth interviews was started on November 29, 2001 and was completed by December 18, 2001. The questions for the first interview with administrators are attached in Appendix D. The second round of follow-up informal interviews was completed May 1, 2002. The initial interviews began with a brief discussion of the research study that was being completed and a review of the interview protocol to be used. Permission to tape record the interviews, with an assurance of confidentiality, was sought and obtained. The typical principal interview lasted 60 minutes.

One set of semi-structured interviews was also conducted with three teachers from each building. The interviews included two teachers from different teaching assignments (i.e. one primary and one intermediate) selected by the principal and a third teacher who was also a building association representative. The Iowa State Education Association (ISEA) provided a list of appropriate building contacts who were also association representatives. When these individuals were available, they were included in the interview triad. The intent was to solicit responses from typical individuals so no attempt was planned to identify individuals considered extreme or deviant in their attitudes or role performance. While this was a compromise from the ideal of a random sample, there is no reason to believe that the teachers were selected according to demographic or attitudinal characteristics that would lead to biased results. The same researcher

conducted all interviews. The teacher interviews began with a brief discussion of the research study that was being completed and a review of the interview protocol to be used. Permission to tape record the interviews, with an assurance of confidentiality, was sought and obtained from the teachers. To further protect confidentiality, demographic and other information was not obtained. This will prohibit a more detailed description of the teacher sample, but was chosen to encourage free and open responses to the interview questions. The typical teacher interview lasted 30 minutes.

Surveys

In order to provide for a wider range of participation, a 4-point Likert scale survey was developed and administered in April 2002 (see Appendix E). Survey items were constructed to assess the teachers' perceived effects of the variable pay on the participants, on the school climate, and on student achievement. Some items collected from over 100 items from prior studies on employee reactions to TBVP were adapted and other new items were written. Items were rated from "disagree" (1) to "agree" (4). An accompanying cover letter explained the purpose of the survey. The survey was distributed to all staff included in the pay plan for the building along with an addressed, stamped envelope addressed to the Iowa Department of Education in the Grimes State Office Building.

Observations

At least one observation of staff professional development in each building was completed. A format for the observation was developed. Field notes of each observation were taken to aid in the examination of the staff development program in the building as to design, opportunities for collaboration or follow-up, proximity to the classroom, and relationship to building goals.

Informal observations were also completed as the opportunity arose. These included staff interactions, building climate, classrooms, and staff areas. Again field notes were recorded that included the people, events, or situation involved, the main issues in the contact, and the implications on this study.

Artifacts

Building demographic information was accumulated from the Basic Educational Data Survey (BEDS) documents and building websites. District goal information and test data were collected from the Comprehensive School Improvement Plan (CSIP) and the Annual Progress Reports (APR) for 2000-2001 and 2001-2001. Criterion-referenced tests (CRT) data and norm-referenced test (i.e. Iowa Tests, ITBS and ITED) data was collected for at least the last two years for each building. Information was gleaned from additional documents including the school's application for the pilot program and district documents.

Analysis

As an embedded case study, there were two units of analysis to be examined. At the finer grain of analysis, case studies were written for each school (Appendix B). In turn, these cases

informed the larger unit of analysis. The eighteen cases were compared to create a cross case analysis, which form the main portion of these results.

The taped interviews both with the principal and with the teachers were transcribed to facilitate a content analysis of responses. The analysis sought to identify major themes or issues embedded in the principals' and teachers' responses to the questions. Standard qualitative methods of content analysis were employed.

The surveys were quantitatively tabulated with any additional comments noted. The units of the analysis were the individual staff member and the individual school. The means, standard deviations, and correlations of teachers' agreement ratings for the outcomes were computed (Appendix E and F). Factor analysis of the intercorrelations was conducted using SPSS software to determine which outcomes clustered together. Means, standard deviations, and intercorrelations among the variables were completed

To facilitate the management of the data an unordered meta-matrix (Merriam, 1988), that is a large chart organized by key variables, was designed to include key phrases, quotes, and other illustrations of a category. A coding system evolved from this. Data management was facilitated through the use of technology. The constant comparative method provided by Glaser and Strauss (1967) and adapted by Lincoln and Guba (1985) was undertaken. The synthesis of the data included identification of themes, ideas, and relationships. Hierarchies of concepts and categories and explanations of concepts were established when possible. Triangulation of information was completed for each concept established. Member checking was accomplished with principals and the interviewed teachers of the 18 schools. It should be noted that the teacher statements are based on limited samples and should be treated as suggestive and tentative rather than necessarily representative of all teachers in the schools.

Quantitative techniques were employed in the area of student achievement. Cohort growth using ITBS or ITED scores in the areas of reading comprehension and mathematics were calculated when the information was available.

Findings

The 18 schools chosen for the pilot project vary widely in many ways (Table 3 and Table 4). For example, the average teacher salary ranges from $29,740 in one school to $42,644 in another. Average teacher experience varies from 8 years in one district to more than 20 years in another. On many characteristics, the schools are representative of the range of values seen in schools in Iowa.

Table 4
Teacher Data from the Pilot Schools*

District, School	Average FT Teacher Salary	Average FT Teacher Total Experience	Average FT Teacher District Experience	Average Age of FT Teacher.	No. FT Teachers with Adv. Degrees	% FT Teachers with Adv. Degrees
Davis County, Davis ES	$29,740	16.7	15.5	42.4	8	20.0
Davis County, Davis MS	$32,159	19.1	17.2	45.1	4	18.2
Des Moines, Oak Park ES	$39,766	10.0	7.4	40.3	11	40.7
Griswold, Elliott ES	$32,404	10.0	9.7	36.6	0	0.0
Johnston, High School	$38,017	11.1	6.9	36.7	14	20.6
Johnston, Lawson ES	$35,464	9.7	5.6	35.4	9	25.7
Johnston, Middle School	$37,919	12.1	7.6	38.9	11	17.2
Linn-Mar, Indian Creek ES	$39,224	15.4	11.4	41.2	7	25.9
Missouri Valley, Middle	$37,632	15.1	13.8	41.9	3	20.0
Northeast Hamilton, K-12	$30,391	8.0	5.3	34.4	3	12.0
Oelwein, Harlen ES	$37,092	13.4	12.9	41.8	2	22.2
Oelwein, High School	$42,102	18.3	15.3	44.4	8	22.9
Oelwein, Middle School	$42,644	20.2	16.6	48.4	5	18.5
Oelwein, Parkside ES	$40,535	17.3	13.7	42	4	44.4
Oelwein, Wings Park ES	$38,121	17.1	14.7	43.9	3	10.3
Stratford, ES	$32,308	9.67	7.56	38.6	0	0.0
Van Buren, Douds ES	$31,776	15.9	13.9	42.3	1	10.0
Van Buren, Stockport ES	$31,942	10.3	9.83	39.3	2	16.7
Unweighted Means	$36,069	13.9	11.4	40.8	5.3	19.2

*Source: Iowa Department of Education, Basic Educational Data Survey, 2001-2002 Staff file.

Nine schools earned awards (Table 5). Among these were 8 of 11 elementary schools, 1 of 4 middle schools, 0 of 2 high schools, and 0 of 1 K-12 district. One high school completed their posttest in November 2002 with results expected in December.

Table 5
Schools Earning Awards

District	Building	Goals met	Goals not met	Reward earned
Davis County	Elementary		X	
Davis County	Middle School		X	
Des Moines	Oak Park ES	X		$41,100
Griswold	Elliott ES		X	
Johnston	Lawson ES	X		$44,400
Johnston	Middle School		X	
Johnston	High School	*		
Linn Mar	Indian Creek ES	X		$43,000
Missouri Valley	Middle School		X	
Northeast Hamilton	K-12		X	
Oelwein	Wings Park ES	X(75%)		$29,550
Oelwein	Harlan ES		X	
Oelwein	Parkside ES	X(75%)		$8,400
Oelwein	Middle School	X(75%)		$28,575
Oelwein	Senior High		X	
Stratford	Stratford ES	X		$9,900
Van Buren	Stockport ES	X		$16,800
Van Buren	Douds ES	X		$15,600
TOTALS		9	8	$237,325

*Fall 2002 testing.

Benefits of TBVP

Some administrators and teachers saw the influence of TBVP as being overwhelmingly positive. Others were much less enthusiastic about the program. This section highlights the major themes and issues found in the comments made by teachers and administrators regarding TBVP.

Goals

During the interviews, staff members in ten schools shared that they felt that TBVP brought the school goals to the forefront. Administrators noted that the awareness of the staff was raised about the school's annual goals. The goals were integrated into the work of each school and the goals provided focus for the school. Staff members discussed the motivating effects of the goals in terms of the pretest and posttest. The difference in the students' scores at the beginning of the year and the end of the year was a source of pride for teachers. Teachers also noted that they were intrinsically rewarded for focusing on the goals whether or not they receive a monetary reward.

A staff member at one school noted that "everyone in the school, students and staff, alike have been made more aware of the goals for the year. The focus is on what is good for the kids." Seeing the goal in black and white and knowing where they want to be at the end of the year was considered motivating. In many of the schools the goals were posted in the rooms of the school building. One building even lined the corridor with student written academic and personal goals. Another noted that TBVP gave them an extra boost of awareness with no extra pressure placed on the students. Awareness was raised and conversations took place about the goals.

Members of one school staff noted that they particularly liked that the focus on goals for all students. Team members in many of the schools appreciated that the goals were in reading. "The better the student reads and understands the better they will do in my class," was what one teacher explained. However, not everyone agreed. One teacher from a high school building stated that collaboration was important in a school building and she felt that this was being undermined by the focus on the core goals.

Teamwork

During the interviews in 14 of the 18 pilot schools staff members highlighted the benefit of teamwork and collaboration as a byproduct of TBVP. One team reported working together to make sure that "no kids fall through the cracks." One staff reported spending more time analyzing what children were learning and discussing how to help students who were not doing well. Other school building teams also reported more discussions by staff on what techniques are working and what is not working to improve student achievement. Teachers in an elementary school expressed the feeling that the staff of their school was pulling together with no one left out. They capitalized upon opportunities to discuss apprehensions and successes. The group stated that even if TBVP were no longer funded, they would continue to work as a team. The team members mentioned that it takes work to facilitate collaboration, but it is worth it and has produced a better learning environment for the students. The burden of improving student achievement is seen as a shared responsibility rather than just the responsibility of the classroom teacher. Another teacher stated that the staff sees itself as supportive of each other. Teachers have increased the sharing of successes with each other and with the students. Even in a school with a staff that considered itself as "pretty focused on achievement of all children before committing to the pilot project" stated that the main change that they have seen is staff working together as a school wide team.

The paraprofessional staff members are also included in the team. Teachers reported seeing the associates as having a greater role in improving student achievement. The associates reported feeling like they were a more integral part of the staff and saw themselves as having value for the school. Teachers in a school that included the entire staff in the TBVP rewards proudly reported every employee with no demarcation line was included in the team effort to improve student achievement. One teacher discussed the cafeteria manager reading with the students, mentors reading with the students, custodians encouraging students to work hard, and the participation of the Parent Teacher Association. In one school a staff member noted that he liked the fact that the financial benefits were extended to all teachers and associates which demonstrated that all staff members work for the children.

Teachers in another school noted that unity has improved. "Everyone has come together as a team to make every student successful." This included more involvement of noncore teachers and auxiliary staff including the custodians. "Everyone is pitching in to help wherever they are needed," a teacher stated. Physical Education teachers in one school offered to help with small reading groups. One teacher who stated that she felt affirmed by the participation in TBVP said, "I like being part of a team where student achievement is praised and recognized by all staff, certified and noncertified.

One principal noted that the teachers were taking greater ownership of the assessment data. A member of the same staff noted that "our staff has really pulled together as a team with excellent leadership from our principal. Students are the winners when we all work together."

Collegiality among faculty members increased in some of the schools. As one teacher stated, "I feel it has brought our staff together and encouraged sharing of ideas and materials that may otherwise not have been shared." Another agreed, "I like what it did to pull this school together." Team members in other schools discussed an increase in communication between staff members.

Rewards

Recognition for doing a good job was the number one benefit discussed by the teachers in one school. Some teachers stated that they liked a reward system based on teachers' hard work toward an achievable goal. Other teachers indicated that the program was comfortable in that it was similar to what they have been doing, but now they will be recognized for their students' achievements. They expressed excitement about this validation. A teacher noted that she "felt like TBVP would reward them (teachers) for being dedicated professionals." She noted that the recognition would also be welcome. Some teachers felt that the extra pay gave them extra incentive and motivation. One noted that "the bonus rewards our hard work. It likely pushes lazy teachers to work harder." Several of the teachers commented that they might have done many of the things that they did this year anyway, but the incentive of TBVP encouraged them to move more quickly to get things done. More than one teacher stated that it was nice to be compensated for their hard work. One response was that "this (TBVP) was a great thing for the state to do. I like getting a bit of a reward for working together for the good of the kids."

Teaching Practice

More than half of the schools discussed changes in curriculum delivery to students. One team reported that their staff was spending more time analyzing what children were learning and discussing how to help students who were not doing well. One teacher noted, "I feel many of the staff members had this pilot in the backs of their minds throughout this year. However, we take great pride in reaching our goals by the action plans we set…this pilot pushed us even more." Teachers in another school discussed increased time spent examining strategies and interventions to see if they are really working. An enhanced focus on student achievement was seen. One staff member noted that the emphasis is on all students improving, not just those at the bottom.

One elementary school reported increased reading activities such as a reading celebration and a continuous read center that have been received enthusiastically by many of the students.

Mathematics activities have also been enhanced with storefront activities and lots of graphing activities. One staff member stated that "if we don't get the money, we don't get the money. It's still win-win because we are helping kids." Teachers at one high school noted that it has not been business as usual. Every department joined in the effort to facilitate improved reading and mathematics skills. One teacher noted that while high schools can tend to be territorial, it is not happening with this project. As the pilot progressed emphasis was less on the possibility of receiving a financial reward and more on the product of improved learning for students. A middle school reported that a conscious effort has been made to integrate reading and mathematics into almost all of the coursework. The teachers noted that their professional development program focusing on reading has helped. They stated that they like the modeling that was going on for kids demonstrating that reading and math skills go beyond the reading and math classroom. They gave the example that no longer were percents seen as only a "math thing" but they are also used in the social science classroom.

The improved use of assessments was also reported. One principal reported that teachers are becoming better at using assessments and probes to group, regroup, and reteach. Teachers also listed assessment literacy among the changes brought by TBVP. In one school teachers reported that they are more likely to examine growth patterns and to chart progress. Some of the teachers felt that they would probably be doing this anyway, but TBVP was an extra push to do it now.

Benefits of the TBVP program were noted at each of the schools involved in the pilot. Staff at one school mentioned that they liked the fact that TBVP was a volunteer program. At another school a teacher stated that she appreciated that this was something that they chose to do with no pressure from administrators or state consultants. In another school teachers noted that the pilot was not "shoved down their throat," but rather they chose to participate. One teacher said that he expects the positives to stay next year even if the funding goes away.

Concerns about TBVP

Disrespect for Professionals

Several concerns about the TBVP program were also noted. Staff members were skeptical of the premise of merit pay. Teachers felt that TBVP did not contribute to the staff being viewed as professional. Apprehensions were raised that they would be viewed as teaching only for the money. One teacher noted, "I have been very insulted about this whole concept. It seems to be the premise that I, as an educator, do not give my utmost in order to have the children in my class succeed, but if I'm given more money, I will value their success more." One commented that "teachers were being reduced to pawns whose actions were dictated by money." Another wrote, "I hope our teacher hopefully give their best without having to rely on 'bonus' pay." Still another stated, "Most of us are in education to help students. We do our best without these incentives." One teacher noted that "the joy of teaching and learning are clouded by this incentives. I wouldn't be teaching if money was my object – bonus or not." One teacher was concerned that she would read in the newspaper that scores went up because the teachers got more money. Teachers were concerned that the public would see teachers as "working for carrots." They wanted to make sure that it was understood that they were doing what they did for the students. Teaching is an important profession and the public perception of the profession is often distorted.

One staff member noted that it is unfortunate in education that we have to bribe teachers to do what they are supposed to do anyway. Teachers in three schools stated that TBVP was another "hoop to jump through" and more politics to play to get extra money. The belief was that teachers should be recognized for what they do and be compensated accordingly without additional requirements.

One team noted that the financial benefits possible were a positive, but were not seen as enough incentive to make poor teachers better. The amount of money involved was not enough to make a difference in motivating teachers. More motivation for teachers comes from high expectations from administrators and a willingness to terminate poor teachers. Similarly another team stated that many of the things they do for students are intangibles, often things that only the teachers, themselves, know about. This team was gratified that they knew that they would have done a lot without the extra pay. One teacher noted concern that this program tended to focus teachers on extrinsic student motivation when what should be emphasized is intrinsic motivation in order to develop students into lifelong learners.

Teachers felt that sufficient funding for teachers and for schools should be provided without a program like TBVP. They were not sure that it was proper for pay to be based this way. Another staff member was concerned that teaching was being diminished to just meeting a goal on a standardized test. Concerns were also noted that TBVP might be just a precursor to an individual merit pay system. One teacher noted that no one in education is in it for the money, but added, "of course, that doesn't' mean that we're so unintelligent we'd turn down extra pay. We all believe teacher salaries are low so this 'incentive' is earned, not a bonus."

A few comments were also noted that considered TBVP in a different way. One teacher mentioned that all teachers are to be rewarded even if they don't participate. She was concerned about whether or not this was really the intent of the legislature. Similarly another staff member reported, "there are only a select few that I feel should benefit from this, most teachers do not deserve this bonus."

Assessments

Another area of concern was the problems caused by making judgements solely on test scores. Test scores are seen as very limited measures. Teachers and administrators in several buildings were concerned that low students would make considerable gains or "tremendous growth", but because they were still below grade level the gains would not be recognized because the students would still not score in the proficient level. A teacher in a middle school was apprehensive about using the ITBS as a measure since it is only one day and one test. "We tried very hard to increase student achievement, but the test is a picture of just one day. I hope it was a good day." One staff member stated that the goals were risky because they could be determined by the way the students' approach the standardized test. The concern of "one day, one test" was also raised in another school where the measure was a criterion-referenced test. Teachers were concerned that the students may not understand or appreciate the importance of the exams. Additional concerns were also discussed about basing the criteria on the Iowa Tests. "If we want all students to improve, an assessment that doesn't measure students against each other is required." Teachers in this school were also concerned that the measurement of their reaching their goals

would be determined by the scores of other schools on the ITED as well as their own. A few additional schools also expressed concern about the emphasis on tests. One principal was concerned that the emphasis was on improving reading and mathematics scores rather than on improving reading and math.

Outside Variables

Many school staff members raised concerns about the factors over which teachers have no control. A teacher stated that teachers can control instruction and curriculum, but there are lots of variables that influence the classroom that are beyond their control. Factors noted included the socio-economic mix of their classroom, the academic mix of their classroom, the students' attitudes about tests, the scheduling of the test, number of special education students included in the classroom scores, and the class size. One staff member stated that their students were not as motivated as students in Ames or Cedar Falls due to lower socio-economic status and ability levels. Teachers in another school were concerned that other student factors such as their high mobility rate should be taken into consideration. One stated that, "there are too many out of school situations which limit our students' ability to learn or their attitudes toward learning. Additional concerns were articulated about the lack of parent accountability for their student's achievement. "I would like parents to do a better job at raising their kids so I don't have to do it for them." An elementary teacher noted, "teachers deserve much better pay across the board, but they don't wok harder only if they get more money. Evaluate curriculum, methods, classroom management, whatever I can control. Don't base bonuses on what I can't, a student's ability and environment outside the school."

Time

A few schools discussed the use of time as one of the drawbacks to TBVP. In one school teachers were concerned that the time spent on TBVP takes away from something else. Another teacher discussed the increase in the number of meetings due too TBVP. Concerns also centered on the amount of time taken by the documentation of activities to meet the goals that were required by the administration. Teachers in another school expressed concern about the time spent on their work with standards and benchmarks. They said that they needed the time to talk to each other and share activities.

Pressure

Staff members in at least five of the schools relayed concerns about increased pressure felt by the teachers to produce achievement in their students. Staff members reported increased anxiety levels in the teachers this year. Staff in one school reported increased pressure on the teachers that teach reading, mathematics, and science. They were also concerned that the focus on reading, mathematics, and science might alienate the other disciplines.

Proposals

Additional concerns were raised concerning the TBVP building plan in some of the districts. Staff members in one building mentioned that they had rushed into the pilot. In another building

teachers were concerned that the pilot project could have been thought out a bit more if the pilot application had not been so hurried. Teachers interviewed also noted that the state was vague in the TBVP application although they couldn't decide if this was good or bad.

Teachers in one school felt like they were "in the dark" about what the central administration was trying to do. Concerns articulated included an apprehension that one teacher may feel like they have failed the school if they do not show the needed growth. Another staff in a district where more than one building was participating in TBVP was concerned about the dynamics that will be involved if one building doesn't meet the district goals, but the others do. A similar concern was raised that some people will work harder than others do and will still get the same bonus. One teacher noted that it would be easy to let someone else take over and just go along for the ride, but she was quick to add that this was not happening in her building this year. Another staff member took issue with the program, "While I believe teaching salaries should be raised and a merit pay for student achievement would be appropriate, this seems to me to be the worst of both worlds: variable pay without individual accountability." One school also had some concerns about the division of associates who were eligible for bonuses and those who were not. In their school only the associates who work directly with students were included in the plan.

Funding

In at least five of the schools concerns were articulated about the problems of funding of the TBVP. Some teachers expressed displeasure that the TBVP program would not be funded next year. They voiced the concern that other schools should also be able to reap the rewards. One teacher stated displeasure with the state "starting programs, but not being able to fully fund them from year to year." Funding resources for next year within the building were also a concern. Some staff interviewed stated that they were concerned that their position would not continue for the next school year.

Nothing Different

Teachers in five of the schools remarked that they were not doing any different because of TBVP. One staff member stated that since they were setting goals anyway they decided that they might as well get something out of it. Another teacher noted no advantages or disadvantages. She felt that the school is continuing to do what it has always done with no changes having taken place due to TBVP.

Student Achievement

Although schools used either ITBS or curriculum-based measures (CBM), the goals varied considerably from district to district (Table 6). With everything else equal, school buildings that based their growth on average growth in the building or the percent of students moving from basic to proficiency in the building gave themselves goals to meet that had less variation than schools that based their goals on classroom or grade level performance. School buildings that based their goals on growth in each grade failed to meet their goals when one grade did not grow at the predicted rate. The Davis County and Oelwein Districts used grade level results, but allowed awards when most, not necessarily all, grades met the goals.

Table 6
Summary of the Approach Buildings Used to Set Goals

District	Building	Goals met	Assessment used				Goals		Partial awards
			ITBS	CBM	NWEA	SIMIII	Bldg	Grade	
Davis County	Elementary	N		X				X	Y
Davis County	Middle School	N		X				X	Y
Des Moines	Oak Park ES	Y		X			X		N
Griswold	Elliott ES	N	X	X	X			X	N
Johnston	Lawson ES	Y	X				X		N
Johnston	Middle School	N	X				X		N
Johnston	High School	*	X				X		N
Linn Mar	Indian Creek ES	Y	X	X			X		N
Oelwein	Harlan ES	N	X					X	Y
Oelwein	High School	N	X					X	Y
Oelwein	Middle School	Y	X					X	Y
Oelwein	Parkside ES	Y	X					X	Y
Oelwein	Wings Park ES	Y	X					X	Y
Mo. Valley	Middle School	N			X			X	N
NE Hamilton	K-12	N	X			X		X	N
Stratford	Stratford ES	Y		X				X	N
Van Buren	Douds ES	Y	X	X			X		N
Van Buren	Stockport ES	Y	X	X			X		N

*Pending fall testing
**NWEA and SIM III are special CBMs.

The amount of growth required to meet the building goal also varied by school. The Johnston School District has set a long-term goal that they want to reach in five years. By dividing the difference between their current status and that of this long-term goal into five increments, they set yearly goals for the five-year period. Oak Park examined past performance and chose an increment of increase that was greater than they had attained in the past, but that they also believed was attainable. The Oelwein District decided to take an unusual tract as they used sections of the ITBS (Reading Comprehension and Mathematics) as their pretest and posttest. Unlike Johnston and Oak Park, Oelwein decided to not just increase the percent of their students scoring at the proficient level on their assessment, but also to increase the percent of their students who scored at the advanced level.

All schools reported growth for their students. However, whether or not this growth was greater than usual is difficult to ascertain. Even when the growth was determined to be an increase over the usual expected growth, it is still difficult to attribute the increase totally to TBVP. The eighteen schools involved implemented or continued many programs within their walls designed to increase student achievement during 2001-02 including, among others, Every Child Reads, Accelerated Mathematics, and Red Schoolhouse.

Quantitative assessment of student achievement change across the 18 schools was problematic. Because effect sizes could not be computed due to the lack of standard deviations for many of the assessments used, the ITBS was used as the basis of comparison. The proficiency levels reported to the state each September were not used due to concerns about increasing the error through the use of cross sectional groups. In addition, the renorming of the ITBS during the

2000-2001 caused increased variability. ITBS scores for each school can be found in Appendix G.

Teacher Motivation

Teacher motivation as a result of the TBVP program was assessed using a survey. Of the 788 surveys sent to schools, 553 (70%) were returned (Table 8). Of the 553 surveys, 78 failed to list the name of the school in which they are employed. Sixteen did not list either the district or the school in which they are employed. 474 (60%) surveys were valid to be disagregated by school. The results from all surveys were used except when school level data was needed to complete the analysis. Complete results of the survey can be seen in Appendices B and F.

Table 8
Returned Survey Counts of the Pilot Schools

District, School	Surveys Sent	Valid Surveys Returned	Percent Valid Returns
Davis County, Davis ES	49	24	49%
Davis County, Davis Middle	33	18	55%
Des Moines, Oak Park ES	54	29	54%
Griswold, Elliott ES	23	19	83%
Johnston, High School	120	46	38%
Johnston, Lawson ES	64	41	64%
Johnston, Middle School	121	80	66%
Linn-Mar, Indian Creek ES	62	56	90%
Missouri Valley, Middle	28	17	61%
NE Hamilton, K-12	46	19	41%
Oelwein, Harlan ES	13	5	38%
Oelwein, High School	41	26	63%
Oelwein, Middle School	30	17	57%
Oelwein, Parkside ES	7	5	57%
Oelwein, Wings Park ES	36	24	67%
Stratford, ES	14	7	50%
Van Buren, Douds ES	23	21	91%
Van Buren, Stockport ES	24	21	88%
No school listed	NA	78	NA
Total	788	553	70%

Survey results ranging from disagree (1) to agree (4) were averaged by teacher and by school. Surprisingly, very little difference was seen whether the means were calculated individually or weighted by school. The survey results were for the most part consistent with the interview findings. Using a factor analysis, the survey was reduced to four factors: (1) value of the program/leadership, (2) motivation, (3) concerns, and (4) goals. More information about these factors is presented in Appendix F.

Table 9
Mean Scores and Standard Deviations for All Certified Staff by Factor

Factor	Mean	Standard Deviation
Value of the Program	2.80	.317
Leadership	3.17	.770
Motivation	3.21	.465
Concerns	2.31	.614
Goals	3.31	.582

Survey statements about the value of the program as seen by the participants were included in the survey. Seventy-three percent of the teachers agreed with the statement that TBVP has led to a greater focus on achievement in their school (mean 2.96, see Figure 1). Fifty-three percent of the teachers disagreed that they were doing anything differently this year due to TBVP (mean 2.34, see Figure 2). Similar results were found with the statement "I asked more from my students this year" (mean 2.81).

Figure 1

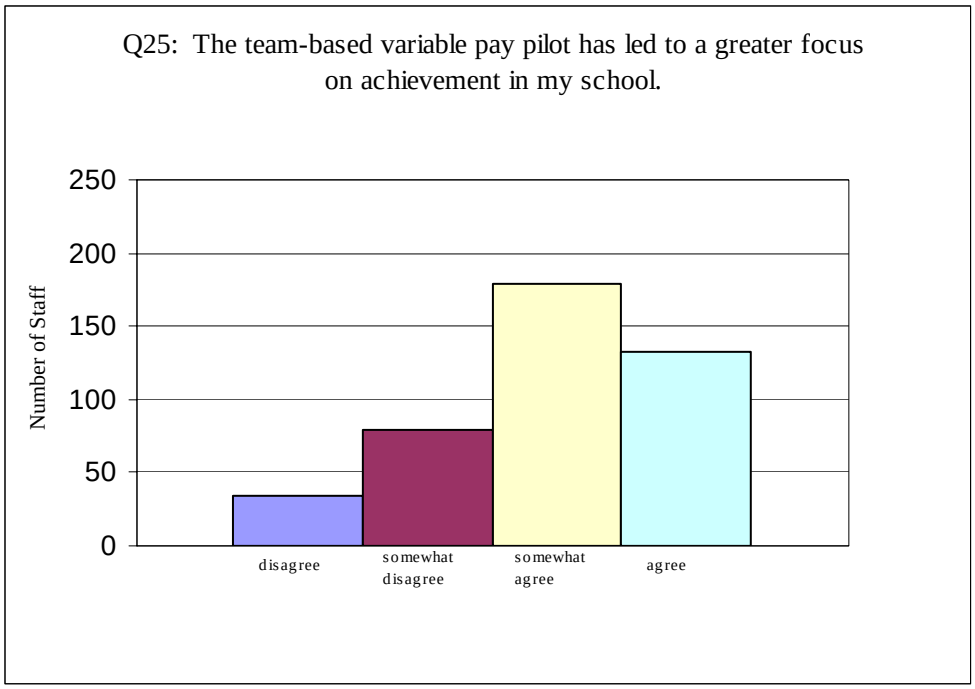
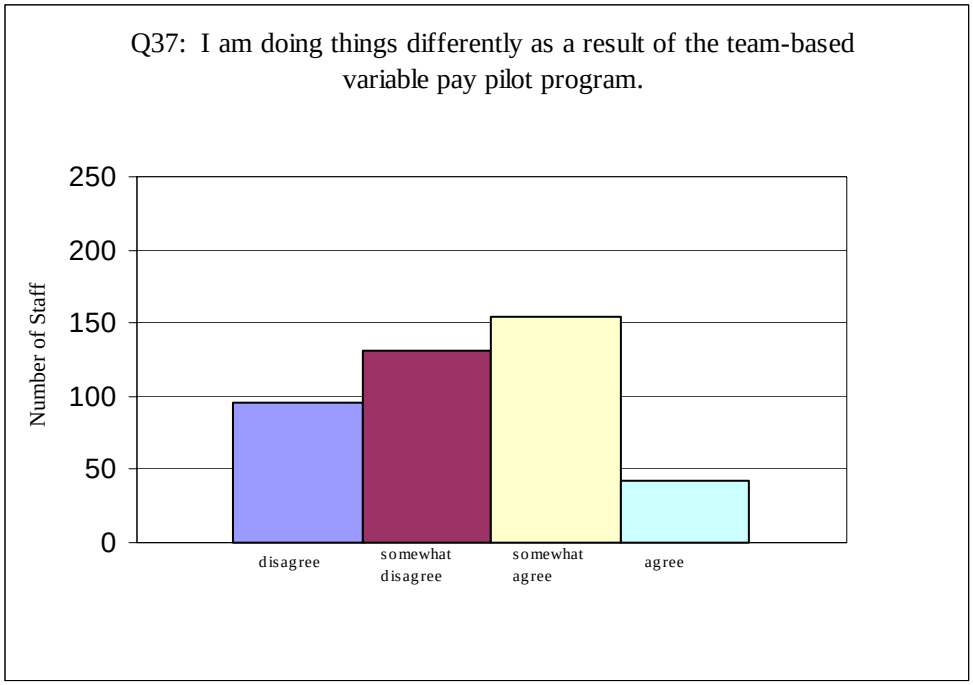


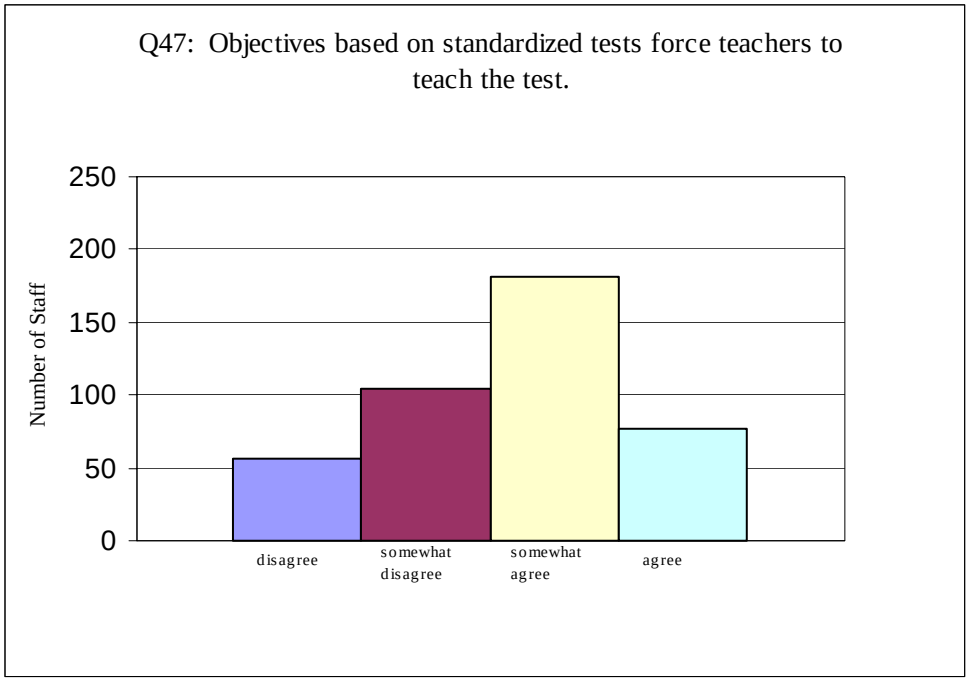
Figure 2



Teachers were asked about the leadership in their schools particularly regarding adequate communication to staff about the bonus process (mean 2.93), support of building principal (mean 3.41), academic leadership of principal (mean 3.21), and work of principal to achieve accountability goals (mean 3.17). Leadership had a large variation in scores among the schools.

Statements about concerns regarding the program received fairly neutral results indicating that about the same number of respondents felt that the concern was present as did not agree that the concern was present. Problems included were concern about the emphasis on testing narrowing the curriculum (mean 2.28), teaching the test (mean 2.66), TBVP requires a lot of extra work (mean 2.23), and TBVP increased stress (mean 2.01). The concern about teaching the test received the most agreement (Figure 3). The concern factor also had a large variation in scores among the pilot schools.

Figure 3



On the survey teachers were asked about their evaluation of their school's goals. On the average, teachers reported strong agreement that their goals were specific (mean 3.56), attainable (mean 3.25), and challenging (mean 3.43). They also agreed that the student achievement goals provide strong focus for their work (mean 3.23).

Two program summary statements were included in the survey (see Figure 3 and Figure 4). The first statement concerned whether or not the staff would work just as hard to achieve the school's goals even without a bonus. Ninety three percent believe that they would work just as hard (mean 3.48, standard deviation .665). The second statement asked for agreement as to whether or not the program should be continued. Eighty seven percent said that the program should be continued (mean 3.39, standard deviation .818).

Figure 4

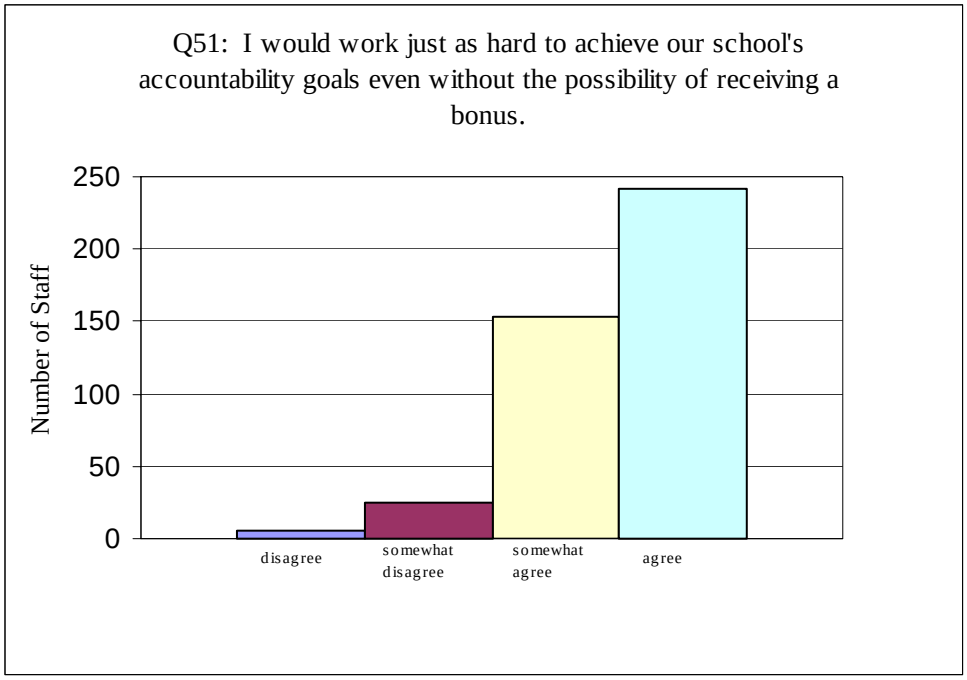
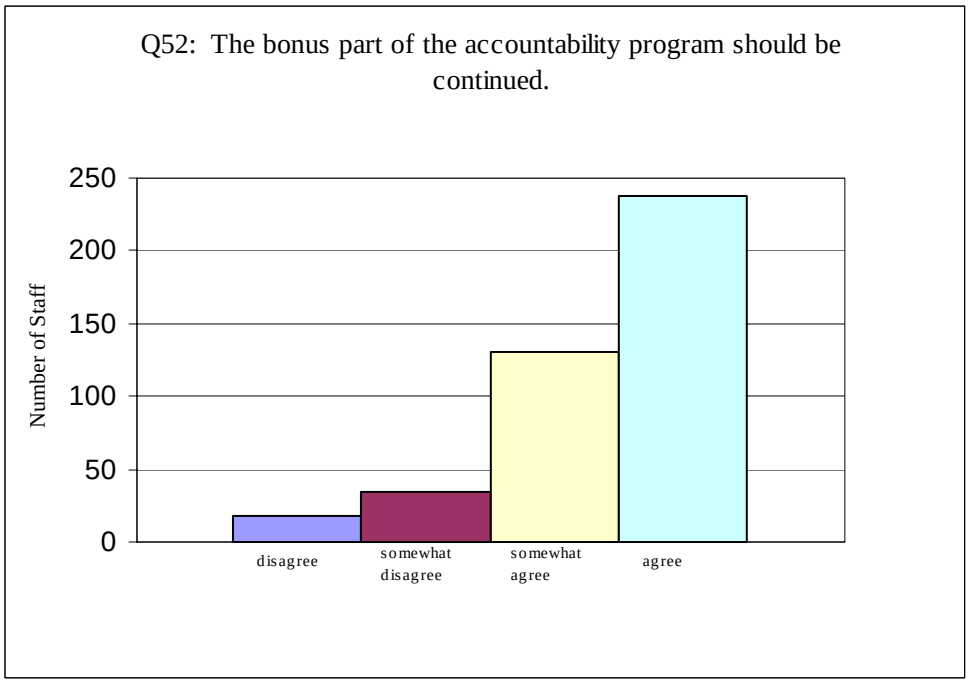


Figure 5



Only two of the survey questions gave significant differences (alpha equal .05) when the schools receiving awards were compared to the schools not receiving awards (Table 10).

Table 10
Means of Survey Statements with Significant Differences between Award Schools and No Award Schools (Scale: 1-Disagree to 4-Agree)

Statements	Award	No Award
Q7: It is appropriate for support staff to receive bonuses.	3.3	2.9
Q23: If the bonus were increased to $3000 my motivation to meet our school's goals would greatly increase.	2.3	3.0

Caution is urged in the interpretation of these differences due to the small number of schools involved. Numerous explanations are possible. Differences in demographic variables (e.g. size, percent minority, percent receiving free or reduce lunch, etc.) between the schools who received the award and those who did not were also checked for any significant differences. Only one significant difference was found. Schools with lower student to staff ratios (mean 11.9) were more likely to receive a bonus than were schools with high student to staff ratios (mean 13.8). However, when only elementary schools are considered, there was no significant difference in student to staff ratios between schools who won awards and those who did not.

Discussion and Conclusion

The results of this study include many inconsistencies. Preliminary results appear that TBVP may increase focus, teamwork, and, perhaps, student achievement in a school building, but the cost may include teacher satisfaction and stress. Certified staff members said in the survey that they would like to see the program continued, but at the same time disagreed as to the value of the program. Most agreed that they would work just as hard to improve student achievement without TBVP. A few schools reported many changes due to TVBP; others reported no change. "Discovering what matters to teachers and how best to motivate them for sustained and improved work is apparently a complicated puzzle, one that is yet to be solved (Johnston, 1986)."

TBVP appears to have the potential to focus teachers and administrators attention on important educational goals in some schools. Teamwork was also appreciably enhanced in certain schools. The motivational value of TBVP was increased with active school leadership. This was similar to what Kelley et. al (2000) found in their study of TBVP programs in Charlotte-Mecklenburg and in Kentucky. Supportive principals generate continuous interest and enthusiasm for a TBVP program. The strengths of TBVP programs were maximized with considerable and ongoing communication among administrators and other staff members. Active support from principals increased the value of the program in school buildings. Engaging staff in the design process also appears to be important. Teachers expressed appreciation at being involved in making the decisions about the application for TBVP.

Student achievement appeared to increase in the TBVP pilot schools during the school year. If this was due to TBVP is not clear. Whether or not this change can be sustained across time is also unclear. More study on the effect of TBVP on student achievement in Iowa and in other states should be considered.

Many teachers were emphatic that TBVP encroached upon their status as professionals. Many teachers were very clear that they did not go into teaching for the money. Many were particularly concerned that the public would see them as doing a better job just to earn a bonus. They felt that they were doing as good a job as they could possibly do already. Evidence was seen that "teachers want to teach" (Lortie, 1975). Time spent on activities not valued by teachers was also noted as a drawback.

Teachers were concerned that they would be measured on the results of a standardized test given on one day to their students. "I hope it is a good day," one stated. They also stated that much of their student's performance is due to factors that they cannot control. Factors from parental involvement to class size to the scheduling of the test were mentioned.

Implications

The following issues came to the forefront through the review of projects in other states and through the Iowa pilot project.

Issue one: Goals appeared to be the most motivating part of TBVP. In Iowa TBVP gave meaning for many staff members to the goal setting that schools have completed in the past.

Performance goals must be challenging yet attainable. Student performance goals may be absolute measures or a relative increase. Schools need be informed in advance the criteria for measurement on which the award will be based. Schools need guidance in choosing goals that are rigorous, but achievable, and schools need help to write better goals. The award should consider schoolwide criteria not on classroom or individual gains. Smaller schools will have increased standard errors, but even the largest schools in the state can have significant variability from one year to the next.

Issue two: The assessment used must be compatible with curriculum standards and benchmarks as well as with teaching strategies. Teachers must have confidence that achievement gains will be measured by the assessment used. The assessment used must be technically adequate. That is, it should be reliable, valid, and fair. In the survey teachers were concerned that objectives based on standardized tests force teachers to teach the test. There is a difference between teaching the objectives covered by a test and teaching the test. Similarly there is sometimes a difference between increasing student learning and increasing performance on an assessment.

Issue three: Enabling conditions must be encouraged. School principals must be instructional leaders not just personnel and building managers. School leadership appeared to be a crucial factor in the acceptance of TBVP and the value of the program as seen by the staff in the Iowa pilot.

Issue four: The program may be voluntary requiring an application or include all schools. By making the pilot program voluntary with schools setting their own goals, increased local ownership was seen. School personnel were able to tailor the program to fit their situation. Teachers in Iowa indicated that they appreciated that this was something that they chose to do.

Issue five: Standards, accountability, and compensation based on performance put additional pressure on teachers. The system must be designed so that the positives of TBVP outweigh the negatives. Teacher stress was more apparent in other states where the consequences were seen has having higher stakes when compared to the Iowa pilot. Also stress appeared to be a greater concern when the teachers felt that this was something imposed upon them, rather than something that they chose to do.

Issue six: The size of the performance award must be large enough to be meaningful to those it is designed to motivate. The pay plan can cover only certified staff or may include noncertified staff and/or support staff. Provide monetary awards to all professional staff at a school to encourage collaboration at the school level. In Iowa, the teachers did not agree that a three-fold increase of the bonus (from approximately $1000 to $3000) would make any difference.

Issue seven: The model needs time for implementation. The program must be sustainable. Teachers were concerned about the continuity of this program in Iowa. This mirrored the concerns seen in other states. Similarly, research to study the impact of this program on student achievement must be long-term.

Limitations

As Iowa decision-makers deal with the need for quality teachers, this work was designed to inform regarding the effects of team-based variable pay on students and staff. The need for such studies is well-documented, Education Commission of the States president Frank Newman (1996) claims: "There is an appalling lack of tough, effective research-based information to help us decide which policies are right for our states and schools" (p. 1). Without such studies, state decision-makers will "not have the tools to focus accountability on student achievement or to stay the course with proven reform strategies" (Education Commission of the States, 1996).

However, there are limitations of this study. Conclusions reached by this study after one year of implementation are very tentative. Results cannot be expected until two to three years after the program is fully implemented. Furthermore, findings should be considered as particular to Iowa. There may be something particular about Iowa or about the schools involved that make generalization of these findings to other situations inadvisable.

The results reflect substantial methodological limitations. Because each school had a unique implementation the synthesis of the individual studies presented "apple and orange" problems. Quantitative meta-analysis methods were not used due to the lack of student level quantitative data. Measures of school performance were also subject to measurement error.

Lack of control or comparisons groups was also a major limitation. By studying only sites that were accepted into the TBVP program this study was limited to the variation that naturally occurred within the schools involved. Use of a comparison group would have allowed additional study in the motivation of teachers by TBVP and the impact of that motivation on school success.

Recommendations

Two recommendations are suggested by this study. First, more study is warranted prior to implementing the TBVP model statewide. An additional two-year pilot study is recommended. This would allow determination of whether or not the student achievement gains are sustainable over time and how the student achievement gains compare to other similar schools who are not participating in TBVP. Furthermore, impact on the staff over time could be considered. The stability of the teacher motivation factors and administration support factors should be studied. Preference for allowing the 18 TBVP sites that are willing to continue in the TBVP program should be given to permit a longitudinal study of three-years for those schools.

Second, along with implementation of TBVP should come technical assistance in the areas of improving goal setting, maximizing professional development opportunities for staff, and assessing students appropriately. Each of these areas appeared to impact the outcomes of the TBVP schools. Technical assistance in these areas is already being planned for all schools through implementation of other sections of the Iowa Teacher Quality Legislation. Specific assistance to TBVP pilot schools would complement the state technical assistance initiatives.

References Cited

Adsit, J., Carpenter, G., and Goff, J. (1998). The building based awards pilot program: An analysis of the results. Jefferson County Public Schools, Golden, CO.

Archer, J. (2000). NEA delegates take hard line against pay for performance. *Education week*, July 12.

Archer, J. (2001). Denver pay plan offers lessons, review says. Education week, December 12, 5.

Becker, W.E. (1998). *Standards and testing: Another view*. Retrieved 2002 from http://www.indiana.edu/.

Bookhart, S, and Freeman, D. (1992). Characteristics of entering teacher candidates. *Review of Educational Research*, 62(1), 37-60.

Boston Public Schools. (2000). City's business, higher ed, non-profit leaders back Boston schools on MCAS, ed reform. Author. Retrieved 2002 from http://boston.k12.ma.us/bps/news_04_14_00.asp.

Burgess, S., Croxson, B., Gregg, P., and Propper, C. (2001). The intricacies of the relationship between pay and performance for teachers: Do teachers respond to performance related pay schemes? University of Bristol, Bristol, United Kingdom.

California Department of Education. (2001). School site employee performance bonus. Author. Retrieved 2002 from http://www.cde.ca.gov/ope/awards/schoolsite/programinfo.htm.

Clark, P.M. and Wilson, J.Q. (1961). Incentive systems: A theory of organizations. *Administrative science quarterly*, 6, 129-166.

Clotfelter, C.T. and Ladd, H.F. (1996). Recognizing and rewarding success in public schools. *In Holding schools accountable: Performance-based reform in education*, H.F. Ladd, ed, 23-63. Washington, D.C.: The Brookings Institution.

Community Training and Assistance Center. (2001). Pathway to results: Pay for performance in Denver. Boston, MA. Retrieved 2002 from http//www.denver.k12.ca.us/pdf/PayForPerformance.pdf.

Consortium for Policy Research in Education. (1995). Linking Compensation to organizational needs. Author. Retrieved 2002 from http://www.ed.gov/pubs/CPRE/fb6/fb6c.html.

Consortium for Policy Research in Education. (1999). Maryland school performance

assessment program. Author, Wisconsin Center for Education Research, University of Wisconsin-Madison. Retrieved 2002 from http//www.wcer.wisc.edu/cpre/teachercomp.

Consortium for Policy Research in Education. (2000). Denver's pay for performance plan. Author. Retrieved 2002 from http://www.wcer.wisc.edu/cpre/tcomp/state/co/denver.asp.

Cooper, S. (2002). Pay factsheet: Variable pay. Employment Relations Unit: Modern Pay Solutions, Employers' Organisation for Local Government, February.

Cooper, S.T. and Cohn, E. (1997). Estimation of a frontier production function for the South Carolina educational process. *Economics of education review*, 16, 313-327.

Council of Chief State School Officers. (1998). State education accountability systems: Kentucky. Author. Retrieved 2002 from http://www.ccsso.org/kyprofile.html.

Council of Chief State School Officers. (1998). State education accountability systems: Maryland. Author. Retrieved 2002 from http://www.ccsso.org/mdprofile.html.

Council of Chief State School Officers. (1999). State education accountability systems: New Jersey. Author. Retrieved 2002 from http://www.ccsso.org/njprofile.html.

Douglas County Schools Group Incentive Board. (2001). *Group incentive program.* Author. Castle Rock, CO: Douglas County Schools.

Education Commission of the States. (2001). Pay-for-performance: Key questions and lessons from five current models. Author. Retrieved 2002 from http://www.ecs.org/clearinghouse/28/30/2830.htm.

Educational Commission of the States. (1996). *The Progress of Educational Reform: 1996.* Denver, CO: Educational Commission of the States.

Elmore, R.F., Abelmann, C.H., and Fuhrman, S.H. (1996). The new accountability in state education reform: From process to performance. *In Holding schools accountable: Performance-based reform in education*, H.F. Ladd, ed, 23-63. Washington, D.C.: The Brookings Institution.

Espinet M., Simmons, P., & Atwater, M. (1992). Career decisions of K-12 science teachers; Factors influencing their decisions and perceptions toward science teaching. *School Science and Mathematics*, 92, 84-91.

Feistritzer, C.E. (1986). *Profile of teachers in the U.S.* Washington, D.C.: National Center for Education Information.

Ferguson, R.F. and Ladd, H.F. (1996). How and shy money matters: An analysis of

Alabama schools. *In Holding schools accountable: Performance-based reform in education*, H.F. Ladd, ed, 23-63. Washington, D.C.: The Brookings Institution.

Florida Department of Education. (2001). Florida school recognition program. Author. Retrieved 2002 from http://www.firn.edu/doe/bin00048/schrmain.htm.

Georgia Department of Education. (2000). The pay for performance program: Evidence of program impact. Author.

Georgia Department of Education. (2001). Pay for performance. Author. Retrieved 2002 from http://www.doe.k12.ga.us/budget/pfp.asp.

Glasser, B.G. and Strauss, A.L. (1967). *The discovery of grounded theory*. Chicago: Aldine.

Goldhaber, D. (2001). Ways that states use assessment data. *CBE basic education online edition*, 45 (6).

Grissmer, D. and Flanagan, A. (1998). Exploring rapid achievement gains in North Carolina and Texas. National Education Goals Panel.

Hajnal, V.J. and Dibski, D.J. (1993). Compensation management: Coherence between organization directions and teacher needs. *Journal of educational administration*, 31, 53-69.

Hall, G. and Caffarella, E. (1998). Executive summary: Third year implementation assessment of the performance pay plan for teachers (1997-'97). Presentation to the Douglas County Board of Education May 5, 1998. Retrieved 2002 from http://www.dcsd.k12.co.us/district/hr/Third.year.assess.98.html

Hambleton, R.K., Jaeger, R.M, Koretz, D., Linn, R.L., Millman, J., Phillips, S.E. (1995). Review of the measurement quality of the Kentucky Instructional Results Information Systems, 1991-1994. A report prepared for the Office of Education Accountability, Kentucky General Assembly, Frankfort, KY: Office of Education Accountability.

Hanushek, E. Kain, J., and Rivkin, S. (1998). Teachers, Schools, and Academic Achievement. NBER Working Paper 6691.

Hartman, D.B. and Weil, R. (1997). Developing a performance pay plan for teachers: A process, not an event. Retrieved 2002 from http://www.aft.org/research/models/dougco/pppwebar.htm.

Heneman, H.G. III. (1998). Assessment of the motivational reactions of teachers to a school-based performance award program. *Journal of personnel evaluation in education*, 12: 43-59.

Heneman, H.G. III. (1999). Teacher attitudes about teacher bonuses under school-based performance award programs. *Journal of personnel evaluation in education,*12(4): 327-341.

Heneman, Herbert III, and Anthony Milanowski. (1999). Teacher Attitudes About Teacher Bonuses Under School-Based Performance Award Programs. *Journal of Personnel Evaluation in Education*, 12 (4) 327-342.

Hoff, D.J. (2001). Test dilemma: Revisions upset trends in data. Education Week, May 2. Retrieved 2002 from http://www.edweek.org/ew/ewstory.cfm?slug=33cutscores.h20.

Hong, N., Al-Khatib, W., Magagna, B., McLoughlin, A. and Coe, B. (2001). *Systems theory*. Unpublished paper. Penn State University, State College, PA. Retrieved 2002 from http://www.ed.psu.edu/insys/ESD/systems/theory/SYSTHEO2.htm.

Huffman, S. and Wilhelmus, M.T. (2000). 838 schools earn state incentive awards of $3.2 million. Retrieved 2002 from http://ideanet.doe.state.in.us/reed/newsr/00January/incent00.html.

Jacobson, S.L. (1999). *Monetary incentives and the reform of teacher compensation: A persistent organizational dilemma*. Unpublished paper. State University of New York at Buffalo. Retrieved 2002 from http://www.gse.buffalo.edu/fas/Jacobson/629/readings/comp2.htm.

Johnson, H., Leak, El, Williamson, G., Kellor, E., Milanowski, T., Odden, A., and Hanna, J. (1999). A case study of the State of North Carolina's school-based performance award program. Consortium for policy research in education, Wisconsin Center for Education Research, University of Wisconsin-Madison. Retrieved 2002 from http//www.wcer.wisc.edu/cpre/teachercomp.

Johnston, S.M. (1986). Incentives for teachers: What motivates, what matters. *Educational administration quarterly*, 22(3), 54-79.

Johnston, R.C. (2000). Louisville school ratings at odds with Kentucky goals. Education week, November 15. Retrieved 2002 from http://www.edweek.org/ew/ewstory.cfm?slug=11louis.h20.

Joyce, B. and Showers, B. (1995). *Student achievement through staff development.* With Plains, NY: Longman.

Just, A.E., Boese, L.E., Burkhardt, R., Carstens, L.J., Devine, M. and Gaffney, T. (2001). Immediate intervention/underperforming schools program (II/USP): How low performing schools in California are facing the challenge of improving student achievement. California Department of Education. Retrieved 2002 from

http://cde.ca.gov/psaa.

Kelley, C. (1999) The motivational impact of school-based performance awards. *Journal of personnel evaluation in education*, 12(4), 309-326.

Kelley, C. (2000). Douglas County Colorado performance pay plan. Consortium for policy research in education, Wisconsin Center for Education Research, University of Wisconsin-Madison. Retrieved 2002 from http//www.wcer.wisc.edu/cpre/teachercomp.

Kelley, C. and Protsik, J. (1997). Risk and reward: Perspectives on the implementation of Kentucky's school-based performance award program. *Educational administration quarterly*, 33(4), 474-505.

Kelley, C., Heneman, H., and Milanowski, A. (2000). *School-based performance award programs, teacher motivation, and school performance: Findings from a study of three programs.* Consortium for Policy Research in Education, University of Pennsylvania.

Kelley, C., Odden, A., Milanowski, A., and Heneman, H. (2000). *The motivational effects of school-based performance awards.* Consortium for Policy Research in Education, University of Pennsylvania.

Kentucky Department of Education. (2001). *Briefing packet, state and regional release, Commonwealth Accountability Testing System (CATS): Accountability 2002 midpoint report*. Author. Louisville, KY: Kentucky Department of Education.

Koehler, W.F. (1996). A K-12 faculty salary-growth model that provides equitable strategic increments. *Journal of personnel evaluation in education*, 10, 37-51.

Koretz, D.M. and Barron, S.I. (1998). *The validity of gains in scores on the Kentucky Instructional Results Information System (KIRIS).* Santa Monica, CA: RAND Corporation.

Ladd, H.F. (1999). The Dallas school accountability and incentive program: An evaluation of its impacts on student outcomes. *Economics of education review*, 18, pp. 1-16.

Lashway, L. (2001). *The new standards and accountability: Will rewards and sanctions motivate America's schools to peak performance?* ERIC Clearinghouse on Educational Management, Eugene, OR.

Lincoln, Y.S. and Guba, E.G. (1985). *Naturalistic inquiry*. Newbury Park, CA: Sage Publications.

Linn, R.L. (1998). *Assessments and accountability. Center for the Study of Evaluation Technical Report 490.* National Center for Research on Evaluation, Standards, and Student Testing (CRESTT), University of California, Los Angeles, CA.

Locke, E.A. (1968). Toward a theory of task motivation and incentive. *Organizational Behavior and Human Performance*, 3, 157-188.

Lortie, D. (1975) *Schoolteacher*. Chicago: University of Chicago Press, p. 175.

Malen, B. (1999). On rewards, punishments, and possibilities: Teacher compensation as an instrument for education reform. *Journal of personnel evaluation in education*, 12 (4): 387-394.

Manzo, K.K. (1998). How to gauge accountability proving ticklish for North Carolina. *Education week*. 4 November. Retrieved 2002 from http://www.edwweek.org.

Maryland State Department of Education. (1999). *School performance recognition awards.* Author. Retrieved 2002 from http://www.msde.state.md.us/fact sheets/fact20.html.

Maryland State Department of Education. (2000). *State honors 248 improving schools with awards, additional funding.* Author. Retrieved 2002 from http://www.msde.state.md.us/pressreleases/2000/october/2000-1027.html.

Menro, R.L. (1998). Student achievement and school and teacher accountability. *Journal of personnel evaluation in education*, 12(3): 257-267.

Meyer, R.H. (2000). Value-added indicators: A powerful tool for evaluating science and mathematics programs and policies. *NISE Brief*, 3(3), 1-8.

Milanowski, A.T. (1999). Measurement error or meaningful change? The consistency of school achievement in two school-based performance award programs. *Journal of personnel evaluation in education*, 12(4): 343-363.

Mohrman, A.M., Mohrman, S. A., and Odden, A. R. (1996). Aligning teacher compensation with systemic school reform: skill-based pay and group-based performance rewards. *Educational evaluation and policy analysis*, 18, 51-71.

Nation's Report Card. (2000). *State profiles*. Retrieved 2002 from http://nces.ed.gov/nationsreportcard/states/.

North Carolina State Department of Public Instruction. (2000). *History of the ABCs program.* Author. Retrieved 2002 from http://www.ncpublicschools.org/abcs/ABCsHist.html.

North Carolina State Department of Public Instruction. (2001). *Evolution of the ABCs.*

Author. Retrieved 2002 from http://www.ncpublicschools.org/Accountability/reporting/ABCSEvolution.html.

Odden, A., Kellor, E., Heneman, H., and Milanowski, A. (1999). School-based performance award programs: Design and administration issues synthesized from eight programs. Consortium for Policy Research in Education, Wisconsin Center for Education Research, University of Wisconsin, Madison.

Olson, L. (1998). A question of value: In trying to gauge educational performance fairly, some researchers and policymakers are turning to a 'value added' approach. *Education week*, May.

Pennsylvania Department of Education. (2000). 1999-2000 Pennsylvania school performance funding program. Author. Retrieved 2002 from http://www.pde.psu.edu/.

Petrosko, J.M. (2000). Assessment and accountability. Louisville, KY: University of Louisville. Retrieved 2002 from http://www.kier.org/assessment.html.

Popham, W.J. (1999). *Classroom assessment: What teachers need to know*. Boston, MA: Allyn and Bacon.

Public Schools of North Carolina, Division of Accountability Services. (2000). Setting annual growth standards: "The formula." Author, *Accountability Brief*, 1(2), 1-4.

Rogers, L., Bond, S., and Nottingham, J. (1997). *Motivation as a factor in the professional development of preservice science teachers*. Paper presented at the National Association for Research in Science Teaching, Oak Brook, IL.

Sanders, W.L. and Horn, S. P. (1998). Research findings from the Tennessee value-added assessment system (TVAAS) database: Implications for educational evaluation and research. *Journal of personnel evaluation in education*, 12(3), 247-256.

Sanders, W.L. and Rivers, J.C. (1996). *Cumulative and residual effects of teachers on future student academic achievement*. Knoxville, TN: University of Tennessee Value-Added Research and Assessment Center.

Schalock, M.D. (1998). Accountability, student learning, and the preparation and licensure of teachers: Oregon's teacher work sample methodology. *Journal of personnel evaluation in education*, 12(3): 269-285.

Schwedel, A., Veysey, P., Conti, E., Kellor, E. and Odden, A. (2000). A case study of the Boston Public Schools School Improvement Awards. Consortium for Policy Research in Education, Wisconsin Center for Education Research, University of Wisconsin-Madison. Retrieved 2002 from

http//www.wcer.wisc.edu/cpre/teachercomp.

Scott, W.R. (1992). *Organizations: Rational, natural, and open systems*. 3rd ed. Englewood Cliffs, NJ: Prentice Hall.

Serow R. (1994). Called to teach; A study of highly motivated preservice teachers. *Journal of Research and Development in Education*, 27(2), 65-72.

South Carolina Department of Education. 1998. Article 11: Awarding performance. Author. Retrieved 2002 from http://www.lpitr.tate.sc.us/bil97-98/850.htm.

Stevens, N., Spaulding, A., Burleson, C., and Killgore, V. (1998). *Incentive grant programs: A report to the 76th Texas Legislature from the Texas Education Agency*. Texas Education Agency.

Stone, J.E. (2001). Value-added assessment: An accountability revolution. Thomas B. Fordham Foundation. Retrieved 2002 from http://www.edexcellence.net/better/tchrs/16.htm.

Sultanik, J. (2000). Bonus pay for teachers. *American school board journal*, February, 34-36.

Supovitz, J.A. and Poglinco, S.M. (2001). Instructional leadership in a standards-based reform. Consortium for Policy Research in Education. Retrieved 2002 from http://www.cpre.org/Publications/ac-02.pdf.

Triplett, S.E. (1997). *A preliminary analysis of the 1996 National Assessment of Educational Progress (NAEP): The class of 2000*. Research Triangle Park, North Carolina: The Center for Research in Education. (ERIC Document Reproduction Service No. ED 414 304).

Urbanski, A. (1998). Teacher professionalism and teacher accountability: Toward a more genuine teaching profession. *Educational policy*, 12, 449-457.

Vroom, V. (1964). *Work and motivation*. New York: Wiley.

Willis, T., Koch, K., Lampe, G., Young, R., Kellor, E., and Odden, A. (1999). A case study of the state of Kentucky's school-based performance award program. Consortium for Policy Research in Education, Wisconsin Center for Education Research, University of Wisconsin-Madison. Retrieved 2002 from http//www.wcer.wisc.edu/cpre/teachercomp.

Wilms, W.W. and Chapleau, R.R. (1999). The illusion of paying teachers for student performance. Education Week on the Web, Nov. 3. Retrieved 2002 from http://www.edweek.org/ew/ewstory.cfm?slup=10wilms.h19.

Wright, S.P., Horn, S.P., and Sanders, W.L. (1997). Teacher and classroom context effects on student achievement: Implications for teacher evaluation. *Journal of personnel evaluation in education*, 11, 57-67.